Damodar N.
GUJARATI
Dawn C.
PORTER

# ECONOMETRÍA

Quinta edición

McGraw Hill

# Econometría

# Econometría

Quinta edición

**Damodar N. Gujarati**

*Profesor emérito de Economía*
*United States Military Academy, West Point*

**Dawn C. Porter**

*University of Southern California*

**Revisión técnica:**

**Aurora Monroy Alarcón**
*Instituto Tecnológico Autónomo de México (ITAM)*

**José Héctor Cortés Fregoso**
*Centro Universitario de Ciencias Económico-Administrativas (CUCEA)*
*Universidad de Guadalajara*

McGraw Hill

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA
MADRID • NUEVA YORK • SAN JUAN • SANTIAGO • SÃO PAULO
AUCKLAND • LONDRES • MILÁN • MONTREAL • NUEVA DELHI
SAN FRANCISCO • SINGAPUR • SAN LUIS • SIDNEY • TORONTO

**Director Higher Education:** Miguel Ángel Toledo Castellanos
**Editor sponsor:** Jesús Mares Chacón
**Coordinadora editorial:** Marcela I. Rocha M.
**Editor de desarrollo:** Edmundo Carlos Zúñiga Gutiérrez
**Supervisor de producción:** Zeferino García García
**Diseño de portada:** Gemma M. Garita Ramos

**Traductora:** Pilar Carril Villarreal

**ECONOMETRÍA**
**Quinta edición**



Mc Graw Hill Educación


Prolongación Paseo de la Reforma 1015, Torre A,
Piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D. F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

**ISBN: 978-607-15-0294-0**
(ISBN edición anterior: 978-970-10-3971-7)

Traducido de la quinta edición de *Basic econometrics*, by Damodar N. Gujarati, and Dawn C. Porter

0-07-337577-2

0123456789 109786543210

Impreso en México *Printed in Mexico*

# Acerca de los autores

## Damodar N. Gujarati

Después de enseñar durante más de 25 años en la City University of New York y 17 años en el Departamento de Ciencias Sociales de la U.S. Military Academy en West Point, Nueva York, el doctor Gujarati es actualmente profesor emérito de economía de la Academia. El doctor Gujarati recibió el grado de M.Com de la Universidad de Bombay en 1960, el grado de M.B.A. de la Universidad de Chicago en 1963 y el grado de Ph.D. de la Universidad de Chicago en 1965. El doctor Gujarati ha publicado una gran cantidad de trabajos en reconocidas revistas nacionales e internacionales, como *Review of Economics and Statistics*, *Economic Journal*, *Journal of Financial and Quantitative Analysis* y *Journal of Business*. El doctor Gujarati fue miembro del Consejo Editorial de *Journal of Quantitative Economics*, publicación oficial de la Sociedad Econométrica de India. El doctor Gujarati es también autor de *Pensions and the New York Fiscal Crisis* (The American Enterprise Institute, 1978), *Government and Business* (McGraw-Hill, 1984) y *Essentials of Econometrics* (McGraw-Hill, 3a. ed., 2006). Los libros del doctor Gujarati sobre econometría se han traducido a diversos idiomas.

El doctor Gujarati fue profesor visitante de la Universidad de Sheffield, Inglaterra (1970-1971), profesor visitante Fulbright en India (1981-1982), profesor visitante en la Facultad de Administración de la Universidad Nacional de Singapur (1985-1986) y profesor visitante de econometría de la Universidad de Nueva Gales del Sur, Australia (durante el verano de 1988). El doctor Gujarati ha dictado numerosas conferencias sobre temas micro y macroeconómicos en países como Australia, China, Bangladesh, Alemania, India, Israel, Mauricio y la República de Corea del Sur.

## Dawn C. Porter

Dawn Porter ha sido profesora adjunta del Departamento de Administración de Operaciones de la Marshall School of Business de la University of Southern California (USC) desde el otoño de 2006. En la actualidad imparte clases de introducción a la estadística tanto en licenciatura como en maestría en la Escuela de Administración. Antes de incorporarse al cuerpo docente de la USC, de 2001 a 2006, Dawn fue profesora adjunta de la McDonough School of Business en la Georgetown University, y antes de eso fue profesora visitante del Departamento de Psicología de la Graduate School of Arts and Sciences en la New York University (NYU). En NYU impartió diversos cursos sobre métodos estadísticos avanzados y también fue profesora de la Stern School of Business. Obtuvo su doctorado en Estadística en la Stern School.

Las áreas de interés para la investigación de Dawn son análisis categórico, medidas de acuerdo, creación de modelos multivariados y aplicaciones en el campo de la psicología. Su investigación actual examina los modelos de subasta en internet desde una perspectiva estadística. Ha presentado sus estudios de investigación en las conferencias de Joint Statistical Meetings, las reuniones del Decision Sciences Institute, la Conferencia Internacional sobre Sistemas de Información, varias universidades, como la London School of Economics y NYU, así como en diversas series de seminarios sobre comercio electrónico y estadística. Dawn es también coautora de *Essentials of Business Statistics,* 2a. edición, McGraw-Hill/Irwin, 2008. Fuera del ámbito académico, Dawn fue contratada como consultora en estadística de KPMG, Inc. También trabajó como consultora en estadística para muchas otras empresas importantes, entre otras, Ginnie Mae, Inc., Toys R Us Corporation, IBM, Cosmaire, Inc., y New York University (NYU) Medical Center.

Para Joan Gujarati, Diane Gujarati-Chesnut,
Charles Chesnut y mis nietos, “Tommy”
y Laura Chesnut.

**—DNG**

Para Judy, Lee, Brett, Bryan, Amy y Autumn Porter.
Pero muy en especial para mi adorado padre, Terry.

**—DCP**

# Contenido breve

# Contenido

# Prefacio

## Objetivo del libro

La primera edición de *Econometría* se publicó hace treinta años. Con el transcurso del tiempo se registraron avances importantes en la teoría y la práctica de la econometría. En cada una de las ediciones subsiguientes traté de incorporar los principales adelantos en el campo. La quinta edición continúa con esta tradición.

Sin embargo, lo que no ha cambiado a lo largo de todos estos años es mi firme convicción de que la econometría puede enseñarse al principiante de manera intuitiva e informativa sin recurrir al álgebra matricial, el cálculo o la estadística, más allá de un nivel elemental. Parte del material es inherentemente técnico. En ese caso, lo coloqué en el apéndice correspondiente o remito al lector a las fuentes apropiadas. Incluso entonces, traté de simplificar el material técnico para que el lector pueda comprenderlo de manera intuitiva.

La longevidad de este libro ha sido para mí una sorpresa muy grata, al igual que el hecho de que no sólo los estudiantes de economía y finanzas lo usan comúnmente, sino también los estudiantes e investigadores de otras disciplinas, como ciencias políticas, relaciones internacionales, agronomía y ciencias de la salud. La nueva edición, con la ampliación de los temas y las aplicaciones concretas que presenta, será muy útil para todos estos estudiantes. En esta edición dediqué todavía más atención a la pertinencia y oportunidad de los datos reales en el texto. De hecho, agregué unos quince ejemplos ilustrativos y más de treinta ejercicios al final de los capítulos. Además, actualicé los datos de aproximadamente dos docenas de ejemplos y más de veinte ejercicios de la edición anterior.

Aunque me encuentro en la octava década de mi vida, no he perdido mi amor por la econometría, y me esfuerzo por mantenerme al tanto de los principales avances en el campo. Para ayudarme en este empeño, me complace mucho contar ahora con la doctora Dawn Porter, profesora adjunta de estadística de la Marshall School of Business de la University of Southern California, en Los Ángeles, como coautora. Ambos trabajamos mucho para llevar a buen término la quinta edición de *Econometría.*

## Características principales de la quinta edición

Antes de explicar los cambios específicos en diversos capítulos, vale la pena destacar las siguientes características de la nueva edición:

1. Se actualizaron prácticamente todos los datos de los ejemplos ilustrativos.
2. Se agregaron varios ejemplos.
3. En varios capítulos incluimos ejemplos finales que ilustran los puntos tratados en el texto.
4. Se incluyen en el libro listados de computadora relativos a varios ejemplos concretos. La mayoría de estos resultados se basan en **EViews** (versión 6) y **STATA** (versión 10), así como en **MINITAB** (versión 15).
5. Diversos capítulos incluyen varios diagramas y gráficos nuevos.
6. Diversos capítulos incluyen varios ejercicios basados en datos nuevos.
7. Los datos de muestras pequeñas se incluyen en el libro, pero los de muestras grandes están en el sitio web del libro con el propósito de reducir el tamaño del texto. El sitio web también publicará todos los datos del libro, mismos que se actualizarán periódicamente.

8. En algunos capítulos incluimos ejercicios para el aula que requieren que los alumnos obtengan datos por su cuenta y apliquen las distintas técnicas que se explican en el libro. También se incluyen algunas simulaciones Monte Carlo en el libro.

## Cambios específicos de la quinta edición

A continuación se enumeran algunos cambios que se refieren de manera específica a ciertos capítulos:

1. Los supuestos en los que se basa el modelo clásico de regresión lineal (MCRL) que se presentan en el capítulo 3 ahora marcan una distinción cuidadosa entre regresoras fijas (variables explicativas) y regresoras aleatorias. Analizamos la importancia de la distinción.
2. En el apéndice del capítulo 6 se analizan las propiedades de los logaritmos, las transformaciones Box-Cox y varias fórmulas de crecimiento.
3. El capítulo 7 explica ahora no sólo el efecto marginal de una sola regresora sobre la variable dependiente, sino también los efectos de cambios simultáneos de todas las variables explicativas en la variable dependiente. Este capítulo también se reorganizó con la misma estructura que los supuestos del capítulo 3.
4. En el capítulo 11 se presenta una comparación de las diferentes pruebas de heteroscedasticidad.
5. Hay un nuevo análisis del efecto de las *rupturas estructurales* en la autocorrelación en el capítulo 12.
6. Los nuevos temas incluidos en el capítulo 13 son *datos faltantes, término de error no normal* y regresoras *estocásticas,* o *aleatorias.*
7. El modelo de regresión no lineal que se analiza en el capítulo 14 tiene una aplicación concreta de la transformación Box-Cox.
8. El capítulo 15 contiene varios ejemplos nuevos que ilustran el uso de los modelos logit y probit en diversos campos.
9. Revisamos e ilustramos cuidadosamente con varias aplicaciones el capítulo 16 sobre *modelos de regresión con datos en panel.*
10. El capítulo 17 incluye un análisis ampliado de las pruebas de causalidad de Sims y Granger.
11. En el capítulo 21 se presenta un análisis minucioso de las series de tiempo estacionarias y no estacionarias, así como algunos problemas relacionados con varias pruebas de estacionariedad.
12. El capítulo 22 incluye una exposición de razones por las que tomar las primeras diferencias de una serie de tiempo con el propósito de volverla estacionaria puede no ser la estrategia más adecuada en algunas situaciones.

Además de estos cambios específicos, corregimos los errores tipográficos y de otro tipo de ediciones anteriores y simplificamos los análisis de varios temas en los diferentes capítulos.

## Organización y opciones

La extensa cobertura en esta edición proporciona al maestro flexibilidad considerable para elegir los temas apropiados para el público al que se dirige. Aquí se dan algunas sugerencias respecto a cómo podría utilizarse la obra.

**Curso de un semestre para los no especialistas**: Apéndice A, capítulos 1 al 9 y un repaso general de los capítulos 10, 11 y 12 (sin las demostraciones).

**Curso de un semestre para estudiantes de economía**: Apéndice A y los capítulos 1 al 13.

**Curso de dos semestres para estudiantes de economía**: Apéndices A, B y C, y capítulos 1 al 22. Los capítulos 14 y 16 son opcionales. Pueden omitirse algunos apéndices técnicos.

**Estudiantes de maestría y posgrado e investigadores**: Este libro es un útil manual de consulta de los temas principales de la econometría.

## Suplementos

Un sitio web muy completo contiene el siguiente material suplementario:

–Datos del texto, así como datos adicionales de conjuntos grandes a los que se hace referencia en el libro; los autores actualizarán los datos periódicamente.

–Un Manual de soluciones, preparado por Dawn Porter, proporciona las respuestas a todas las preguntas y problemas que se presentan en el texto.

–Una biblioteca de imágenes digitales que contiene todos los gráficos y figuras del texto.

Encontrará más información en www.mhhe.com/gujarati5e. Consulte términos y condiciones con su representante McGraw-Hill más cercano.

# Reconocimientos

Desde la publicación de la primera edición de este libro, en 1978, hemos recibido valiosas sugerencias, comentarios, críticas y consejos de muchas personas. En particular, queremos agradecer la ayuda que recibimos de Michael McAleer, de la Universidad de Western Australia; Peter Kennedy, de la Universidad Simon Frazer en Canadá; así como de Kenneth White, de la Universidad de British Columbia; George K. Zestos, de la Universidad Christopher Newport de Virginia y Paul Offner, de la Universidad Georgetown de Washington, D.C.

También deseamos manifestar nuestro agradecimiento a varias personas que influyeron en nosotros por su erudición. Queremos agradecer especialmente a Arthur Goldberger, de la Universidad de Wisconsin, William Greene, de la Universidad de Nueva York y al finado G. S. Maddala. Seguimos agradecidos con los revisores que aportaron su invaluable conocimiento, críticas y sugerencias a las ediciones anteriores de este texto: Michael A. Grove, de la Universidad de Oregon; Harumi Ito, de la Universidad Brown; Han Kim, de la Universidad de South Dakota; Phanindra V. Wunnava, del Middlebury College y Andrew Paizis, de la City University of New York.

Diversos autores influyeron en la preparación de este texto. En particular, estamos agradecidos con los siguientes: Chandan Mukherjee, director del Centro de Estudios de Desarrollo, de Trivandrum, India; Howard White y Marc Wuyts, del Instituto de Estudios Sociales de Holanda; Badi H. Baltagi, de la Universidad Texas A&M; B. Bhaskara Rao, de la Universidad de Nueva Gales del Sur, Australia; R. Carter Hill, de la Universidad de Louisiana; William E. Griffiths, de la Universidad de Nueva Inglaterra; George G. Judge, de la Universidad de California en Berkeley; Marno Verbeek, del Centro de Estudios Económicos, de KU Leuven; Jeffrey Wooldridge, de la Universidad Estatal de Michigan; Kerry Patterson, de la Universidad de Reading, Inglaterra; Francis X. Diebold, de la Escuela Wharton, perteneciente a la Universidad de Pensilvania; Wojciech W. Charemza y Derek F. Deadman, de la Universidad de Leicester, Inglaterra, y Gary Koop, de la Universidad de Glasgow.

Varios comentarios y sugerencias muy valiosos que proporcionaron los revisores de la cuarta edición mejoraron en gran medida esta edición. Queremos expresar nuestro agradecimiento a los siguientes:

**Valerie Bencivenga**
*Universidad de Texas, Austin*

**Andrew Economopoulos**
*Ursinus College*

**Eric Eide**
*Universidad Brigham Young*

**Gary Ferrier**
*Universidad de Arkansas, Fayetteville*

**David Garman**
*Universidad Tufts*

**David Harris**
*Benedictine College*

**Don Holley**
*Universidad Estatal Boise*

**George Jakubson**
*Universidad de Cornell*

**Bruce Johnson**
*Centre College of Kentucky*

**Duke Kao**
*Universidad de Syracuse*

**Gary Krueger**
*Macalester College*

**Subal Kumbhakar**
*Universidad Binghamton*

**Tae-Hwy Lee**
*Universidad de California, Riverside*

**Solaiman Miah**
*Universidad Estatal de West Virginia*

**Fabio Milani**
*Universidad de California, Irvine*

**Helen Naughton**
*Universidad de Oregon*

**Solomon Smith**
*Universidad Langston*

**Kay Strong**
*Universidad Estatal Bowling Green*

**Derek Tittle**
*Instituto Tecnológico de Georgia*

**Tiemen Woutersen**
*Universidad Johns Hopkins*

Deseamos dar las gracias a los estudiantes y maestros de todo el mundo que no sólo han utilizado este libro, sino que se han comunicado con nosotros en cuanto a diversos aspectos de la obra.

Por su ayuda tras bambalinas en McGraw-Hill, estamos agradecidos con Douglas Reiner, Noelle Fox y Anne Hilbert.

Por último, pero no por eso menos importante, el doctor Gujarati desea dar las gracias a sus hijas, Joan y Diane, por su constante apoyo y aliento en la preparación de ésta y las anteriores ediciones.

*Damodar N. Gujarati*
*Dawn C. Porter*

# Introducción

## I.1 ¿Qué es la econometría?

En términos literales *econometría* significa "medición económica". Sin embargo, si bien es cierto que la medición es una parte importante de la econometría, el alcance de esta disciplina es mucho más amplio, como se deduce de las siguientes citas:

> La econometría, resultado de cierta perspectiva sobre el papel que desempeña la economía, consiste en la aplicación de la estadística matemática a los datos económicos para dar soporte empírico a los modelos construidos por la economía matemática y obtener resultados numéricos.[1]

> . . . la econometría puede definirse como el análisis cuantitativo de fenómenos económicos reales, basados en el desarrollo simultáneo de la teoría y la observación, relacionados mediante métodos apropiados de inferencia.[2]

> La econometría se define como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística se aplican al análisis de los fenómenos económicos.[3]

> La econometría tiene que ver con la determinación empírica de las leyes económicas.[4]

> El arte del econometrista consiste en encontrar un conjunto de supuestos lo bastante específicos y realistas para que le permitan aprovechar de la mejor manera los datos con que cuenta.[5]

> Los econometristas… son una ayuda decisiva en el esfuerzo por disipar la mala imagen pública de la economía (cuantitativa o de otro tipo) considerada como una materia en la cual se abren cajas vacías, suponiendo la existencia de abrelatas, para revelar un contenido que diez economistas interpretarán de 11 maneras diferentes.[6]

> El método de la investigación econométrica busca en esencia una conjunción entre la teoría económica y la medición real, con la teoría y la técnica de la inferencia estadística como puente.[7]

[1] Gerhard Tintner, *Methodology of Mathematical Economics and Econometrics*, The University of Chicago Press, Chicago, 1968, p. 74.

[2] P.A. Samuelson, T.C. Koopmans y J.R.N. Stone, "Report of the Evaluative Committee for *Econometrica*", *Econometrica*, vol. 22, núm. 2, abril de 1954, pp. 141-146.

[3] Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, Nueva York, 1964, p. 1.

[4] H. Theil, *Principles of Econometrics*, John Wiley & Sons, Nueva York, 1971, p. 1.

[5] E. Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1966, p. 514.

[6] Adrian C. Darnell y J. Lynne Evans, *The Limits of Econometrics*, Edward Elgar, Hants, Inglaterra, 1990, p. 54.

[7] T. Haavelmo, "The Probability Approach in Econometrics", suplemento de *Econometrica*, vol. 12, 1944, prefacio, p. iii.

## I.2 ¿Por qué una disciplina aparte?

Como indican las definiciones anteriores, la econometría es una amalgama de teoría económica, economía matemática, estadística económica y estadística matemática. Aun así, la materia merece un estudio separado por las siguientes razones.

La teoría económica hace afirmaciones o formula hipótesis de naturaleza sobre todo cualitativa. Por ejemplo, la teoría microeconómica establece que, si no intervienen otros factores, se espera que la reducción del precio de un bien aumente la cantidad demandada de ese bien. Así, la teoría económica postula una relación negativa o inversa entre el precio y la cantidad demandada de un bien. Pero la teoría por sí sola no proporciona medida numérica alguna de la relación entre los dos; no dice cuánto aumentará o se reducirá la cantidad como resultado de un cambio determinado en el precio del bien. El trabajo del econometrista es proporcionar tales estimaciones numéricas. En otras palabras, la econometría da contenido empírico a gran parte de la teoría económica.

El interés principal de la economía matemática es expresar la teoría económica en una forma matemática (ecuaciones) sin preocuparse por la capacidad de medición o de verificación empírica de la teoría. La econometría, como ya apuntamos, se interesa sobre todo en la verificación empírica de la teoría económica. Como veremos, el econometrista suele emplear ecuaciones matemáticas, propuestas por el economista matemático, pero las expresa de forma que se presten para la prueba empírica. Y esta conversión de ecuaciones matemáticas en ecuaciones econométricas requiere una gran dosis de ingenio y destreza.

La estadística económica se relaciona en primer lugar con la recopilación, procesamiento y presentación de cifras económicas en forma de gráficos y tablas. Éste es el trabajo del estadístico económico, cuya actividad principal consiste en recopilar cifras sobre el producto nacional bruto (PNB), empleo, desempleo, precios, etc. Los datos así reunidos constituyen la materia prima del trabajo econométrico. Pero el estadístico económico no va más allá de la recolección de información, pues no le conciernen las cifras recopiladas para probar las teorías económicas. Sin duda, es el econometrista quien se ocupa de realizar esta labor.

Aunque la estadística matemática proporciona muchas herramientas para esta ciencia, el econometrista a menudo necesita métodos especiales por la naturaleza única de la mayoría de las cifras económicas, pues no se generan como resultado de un experimento controlado. El econometrista, como el meteorólogo, suele depender de cifras que no controla directamente. Como observa Spanos, acertadamente:

> En econometría, el que construye el modelo a menudo se enfrenta a datos **provenientes de la observación** más que de la **experimentación**. Esto tiene dos implicaciones importantes para la creación empírica de modelos en econometría. Primero, se requiere que quien elabore modelos domine muy distintas habilidades en comparación con las que se necesitan para analizar los datos experimentales… Segundo, la separación de quien recopila los datos y el analista exige que quien elabora modelos se familiarice por completo con la naturaleza y la estructura de los datos en cuestión.[8]

## I.3 Metodología de la econometría

¿Cómo proceden los econometristas en el análisis de un problema económico? Es decir, ¿cuál es su metodología? Aunque existen diversas escuelas de pensamiento sobre metodología econométrica, aquí presentaremos la **metodología tradicional** o **clásica**, que aún predomina en la investigación empírica en economía y en las ciencias sociales y del comportamiento.[9]

---

[8] Aris Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, Reino Unido, 1999, p. 21.

[9] Hay un análisis ilustrativo, si bien avanzado, de los métodos econométricos en David F. Hendry, *Dynamic Econometrics*, Oxford University Press, Nueva York, 1995. Véase también Aris Spanos, *op. cit.*

En términos generales, la metodología econométrica tradicional se ajusta a los siguientes lineamientos:

1. Planteamiento de la teoría o de la hipótesis.
2. Especificación del modelo matemático de la teoría.
3. Especificación del modelo econométrico o estadístico de la teoría.
4. Obtención de datos.
5. Estimación de los parámetros del modelo econométrico.
6. Pruebas de hipótesis.
7. Pronóstico o predicción.
8. Utilización del modelo para fines de control o de políticas.

Para ilustrar estos pasos, consideremos la conocida teoría keynesiana de consumo.

## 1. Planteamiento de la teoría o hipótesis

Keynes plantea:

> La ley psicológica fundamental… consiste en que los hombres [y las mujeres], como regla general y en promedio, están dispuestos a incrementar su consumo a medida que aumenta su ingreso, pero no en la misma cuantía del aumento en su ingreso.[10]

En pocas palabras, Keynes postula que la **propensión marginal a consumir (PMC)**, es decir, la tasa de cambio del consumo generado por una unidad (digamos, un dólar) de cambio en el ingreso, es mayor que cero pero menor que uno.

## 2. Especificación del modelo matemático de consumo

A pesar de haber postulado una relación positiva entre el consumo y el ingreso, Keynes no especifica la forma precisa de la relación funcional entre ambas cosas. Por simplicidad, un economista matemático puede proponer la siguiente forma de la función keynesiana de consumo:

$$Y = \beta_1 + \beta_2 X \qquad 0 < \beta_2 < 1 \tag{I.3.1}$$

donde $Y =$ gasto de consumo y $X =$ ingreso, y donde $\beta_1$ y $\beta_2$, conocidos como los **parámetros** del modelo, son, respectivamente, los coeficientes del **intercepto** y de la **pendiente**.

El coeficiente de la pendiente $\beta_2$ mide la PMC. En la figura I.1 se presenta geométricamente la ecuación (I.3.1). Esta ecuación plantea que el consumo está relacionado linealmente con el ingreso, y es un ejemplo de un modelo matemático de la relación entre consumo e ingreso, llamada en economía **función consumo**. Un modelo es simplemente un conjunto de ecuaciones matemáticas. Si el modelo tiene una sola ecuación, como en el ejemplo anterior, se denomina **modelo uniecuacional**, mientras que si tiene más de una ecuación, se conoce como **modelo multiecuacional** (consideraremos más adelante este tipo de modelos).

En la ecuación (I.3.1), la variable que aparece al lado izquierdo del signo de la igualdad se llama **variable dependiente**, y la(s) variable(s) del lado derecho se llama(n) variable(s) **independiente(s)**, o **explicativa(s)**. Así, en la función keynesiana de consumo, la ecuación (I.3.1), el consumo (gasto) es la variable dependiente, y el ingreso, la explicativa.

[10] John Maynard Keynes, *The General Theory of Employment, Interest and Money*, Harcourt Brace Jovanovich, Nueva York, 1936, p. 96.

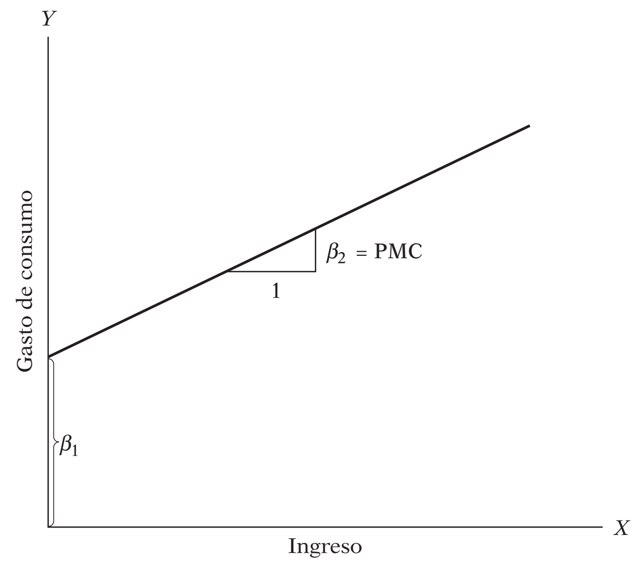



**FIGURA I.1**
Función keynesiana de consumo.

## 3. Especificación del modelo econométrico de consumo

El modelo puramente matemático de la función de consumo dado en la ecuación (I.3.1) es de interés limitado para el econometrista, pues supone una relación *exacta* o *determinista* entre el consumo y el ingreso. Pero las relaciones entre las variables económicas suelen ser inexactas. Así, si fuéramos a obtener información sobre gasto de consumo e ingreso disponible (es decir, después de impuestos) de una muestra de, por ejemplo, 500 familias estadounidenses y graficar estos datos, con el gasto de consumo en el eje vertical y en el eje horizontal el ingreso disponible, no esperaríamos que las 500 observaciones quedaran exactamente sobre la línea recta de la ecuación (I.3.1) porque, además del ingreso, otras variables afectan el gasto de consumo, como el tamaño de la familia, las edades de sus miembros, su religión, etcétera.

Para dar cabida a relaciones inexactas entre las variables económicas, el econometrista modificaría la función determinista de consumo en la ecuación (I.3.1) de la siguiente manera:

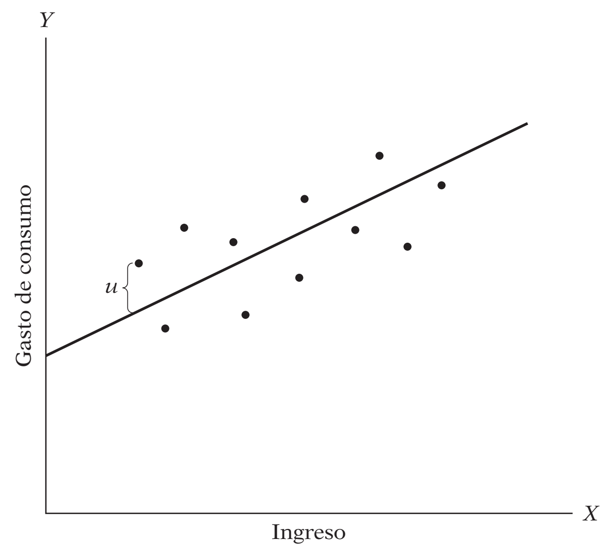
$$Y = \beta_1 + \beta_2 X + u \qquad \textbf{(I.3.2)}$$

donde $u$, conocida como **término de perturbación** o **de error**, es una **variable aleatoria (estocástica)** con propiedades probabilísticas bien definidas. El término de perturbación $u$ representa todos los factores que afectan el consumo pero que no se consideran en el modelo en forma explícita.

La ecuación (I.3.2) es un ejemplo de un **modelo econométrico**. Más técnicamente, dicha ecuación es un ejemplo de un **modelo de regresión lineal**, el principal interés de este libro. La función econométrica de consumo plantea como hipótesis que la variable dependiente $Y$ (consumo) está relacionada linealmente con la variable explicativa $X$ (ingreso), pero que la relación entre las dos no es exacta: está sujeta a variaciones individuales.

El modelo econométrico de la función de consumo se representa gráficamente como aparece en la figura I.2.

**FIGURA I.2**
Modelo econométrico de la función keynesiana de consumo.




## 4. Obtención de información

Para estimar el modelo econométrico dado en la ecuación (I.3.2), esto es, para obtener los valores numéricos de $\beta_1$ y $\beta_2$, son necesarios los datos. Aunque tendremos más que decir en el siguiente capítulo sobre la importancia crucial de los datos para el análisis económico, por el momento observemos unas cifras relacionadas con la economía de Estados Unidos de 1960 a 2005, que se presentan en la tabla I.1. La variable $Y$ en esta tabla es el gasto de consumo personal (GCP) *agregado* (para la economía en su conjunto), y la variable $X$, el producto interno bruto (PIB), una medida del ingreso agregado, ambos medidos en miles de millones de dólares de 2000. Por consiguiente, los datos están en términos "reales", es decir, se midieron en precios constantes (2000). Estos datos se graficaron en la figura I.3 (cf. figura I.2). Por el momento, haga caso omiso de la recta trazada en la figura.

## 5. Estimación del modelo econométrico

Ahora que tenemos los datos, la siguiente labor es estimar los parámetros de la función consumo. La estimación numérica de los parámetros da contenido empírico a la función consumo. En el capítulo 3 explicaremos el mecanismo real para estimar los parámetros. Por el momento, note que la técnica estadística conocida como **análisis de regresión** es la herramienta principal para obtener las estimaciones. Con esta técnica y los datos de la tabla I.1 obtuvimos los siguientes valores estimados de $\beta_1$ y $\beta_2$, a saber, $-299.5913$ y 0.7218. Así, la función consumo estimada es

$$\hat{Y}_t = -299.5913 + 0.7218X_t \qquad \textbf{(I.3.3)}$$

El acento circunflejo (sombrero) sobre $Y$ indica que es un valor estimado.[11] En la figura I.3 se muestra la función consumo estimada (es decir, la línea de regresión).

[11] Por convención, un acento circunflejo (sombrero) sobre una variable o parámetro indica que es un valor estimado.

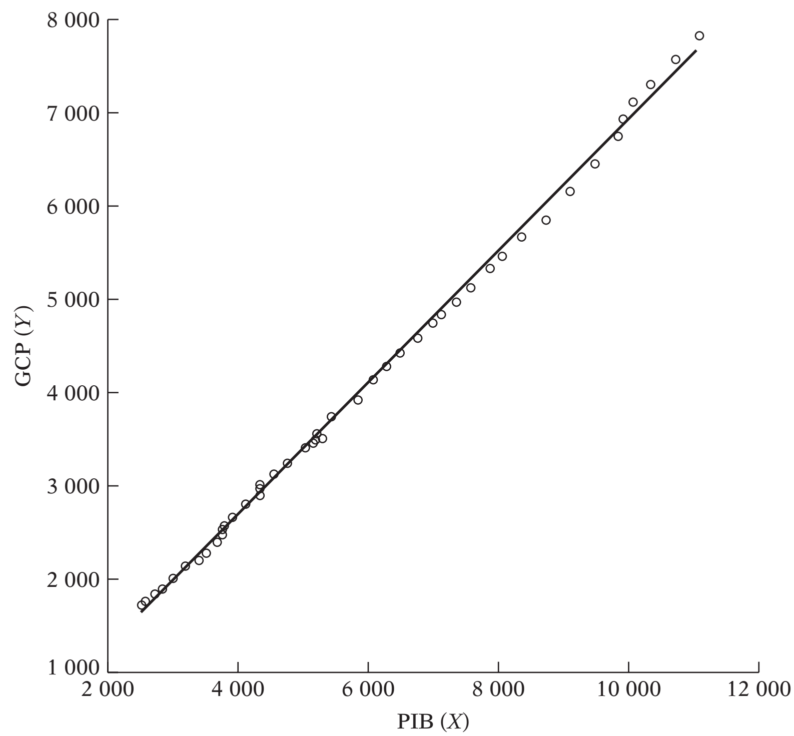
**TABLA I.1**

**Datos sobre *Y* (gasto de consumo personal) y *X* (producto interno bruto, 1960-2005), en miles de millones de dólares de 2000**

Fuente: *Economic Report of the President*, 2007, tabla B-2, p. 230.

| Año | GCP(*Y*) | PIB(*X*) |
|---|---|---|
| 1960 | 1 597.4 | 2 501.8 |
| 1961 | 1 630.3 | 2 560.0 |
| 1962 | 1 711.1 | 2 715.2 |
| 1963 | 1 781.6 | 2 834.0 |
| 1964 | 1 888.4 | 2 998.6 |
| 1965 | 2 007.7 | 3 191.1 |
| 1966 | 2 121.8 | 3 399.1 |
| 1967 | 2 185.0 | 3 484.6 |
| 1968 | 2 310.5 | 3 652.7 |
| 1969 | 2 396.4 | 3 765.4 |
| 1970 | 2 451.9 | 3 771.9 |
| 1971 | 2 545.5 | 3 898.6 |
| 1972 | 2 701.3 | 4 105.0 |
| 1973 | 2 833.8 | 4 341.5 |
| 1974 | 2 812.3 | 4 319.6 |
| 1975 | 2 876.9 | 4 311.2 |
| 1976 | 3 035.5 | 4 540.9 |
| 1977 | 3 164.1 | 4 750.5 |
| 1978 | 3 303.1 | 5 015.0 |
| 1979 | 3 383.4 | 5 173.4 |
| 1980 | 3 374.1 | 5 161.7 |
| 1981 | 3 422.2 | 5 291.7 |
| 1982 | 3 470.3 | 5 189.3 |
| 1983 | 3 668.6 | 5 423.8 |
| 1984 | 3 863.3 | 5 813.6 |
| 1985 | 4 064.0 | 6 053.7 |
| 1986 | 4 228.9 | 6 263.6 |
| 1987 | 4 369.8 | 6 475.1 |
| 1988 | 4 546.9 | 6 742.7 |
| 1989 | 4 675.0 | 6 981.4 |
| 1990 | 4 770.3 | 7 112.5 |
| 1991 | 4 778.4 | 7 100.5 |
| 1992 | 4 934.8 | 7 336.6 |
| 1993 | 5 099.8 | 7 532.7 |
| 1994 | 5 290.7 | 7 835.5 |
| 1995 | 5 433.5 | 8 031.7 |
| 1996 | 5 619.4 | 8 328.9 |
| 1997 | 5 831.8 | 8 703.5 |
| 1998 | 6 125.8 | 9 066.9 |
| 1999 | 6 438.6 | 9 470.3 |
| 2000 | 6 739.4 | 9 817.0 |
| 2001 | 6 910.4 | 9 890.7 |
| 2002 | 7 099.3 | 10 048.8 |
| 2003 | 7 295.3 | 10 301.0 |
| 2004 | 7 577.1 | 10 703.5 |
| 2005 | 7 841.2 | 11 048.6 |

**FIGURA I.3**
Gasto de consumo personal ($Y$) en relación con el PIB ($X$), 1960-2005, en miles de millones de dólares de 2000.




Como se aprecia en la figura I.3, la línea de regresión se ajusta bien a los datos, pues los puntos que corresponden a los datos están muy cercanos a ella. En esta gráfica vemos que de 1960 a 2005 el coeficiente de la pendiente (es decir, la **PMC**) fue de alrededor de 0.72, lo que indica que para el periodo muestral un incremento de un dólar en el ingreso real produjo, *en promedio*, un incremento cercano a 72 centavos en el gasto de consumo real.[12] Decimos "en promedio" porque la relación entre consumo e ingreso es inexacta; como se deduce de la figura I.3, no todos los puntos correspondientes a los datos están exactamente en la recta de regresión. Con palabras sencillas, podemos decir que, de acuerdo con los datos, el *promedio* o *media* del gasto de consumo aumentó alrededor de 72 centavos por cada dólar de incremento en el ingreso real.

## 6. Pruebas de hipótesis

En el supuesto de que el modelo ajustado sea una aproximación razonablemente buena de la realidad, tenemos que establecer criterios apropiados para comprobar si los valores estimados obtenidos en una ecuación como la (I.3.3), por ejemplo, concuerdan con las expectativas de la teoría que estamos probando. De acuerdo con los economistas "positivos", como Milton Friedman, una teoría o hipótesis no verificable mediante la evidencia empírica no puede ser admisible como parte de la investigación científica.[13]

Como ya señalamos, Keynes esperaba que la PMC fuera positiva pero menor que 1. En el ejemplo observamos que la PMC es alrededor de 0.72. Pero antes de aceptar este resultado como confirmación de la teoría keynesiana de consumo, debemos averiguar si esta estimación está lo

[12] No se preocupe aquí por la forma como se obtuvieron estos valores; como veremos en el capítulo 3, el método estadístico de **mínimos cuadrados** produjo estos valores estimados. Asimismo, por el momento no se preocupe por el valor negativo del intercepto.

[13] Véase Milton Friedman, "The Methodology of Positive Economics", Essays in Positive Economics, University of Chicago Press, Chicago, 1953.

bastante abajo de la unidad para convencernos de que no se trata de un suceso debido al azar o de una peculiaridad de los datos. En otras palabras, ¿es 0.72 *estadísticamente menor que 1*? Si lo es, puede apoyar la teoría de Keynes.

Tal confirmación o refutación de las teorías económicas con fundamento en la evidencia muestral se basa en una rama de la teoría estadística conocida como **inferencia estadística (pruebas de hipótesis)**. A lo largo de este libro veremos cómo realizar en la práctica este proceso de inferencia.

## 7. Pronóstico o predicción

Si el modelo escogido no refuta la hipótesis o la teoría en consideración, servirá para predecir el (los) valor(es) futuro(s) de la **variable** dependiente *Y*, o **de pronóstico**, con base en el (los) valor(es) futuro(s) conocido(s) o esperado(s) de la **variable** explicativa, o **predictora**, *X*.

Para ilustrarlo, suponga que queremos predecir la media del gasto de consumo para 2006. El valor del PIB para 2006 fue de 11 319.4 millones de dólares.[14] Colocamos esta cifra del PIB en el lado derecho de la ecuación (I.3.3) y obtenemos:

$$\begin{aligned}\hat{Y}_{2006} &= -299.5913 + 0.7218\,(11\,319.4)\\ &= 7\,870.7516\end{aligned} \tag{I.3.4}$$

o casi 7 870 millones de dólares. Por tanto, con ese valor del PIB, la media o el promedio del gasto de consumo previsto es de alrededor de 7 870 millones de dólares. El valor real del gasto de consumo registrado en 2006 fue de 8 044 millones de dólares. El modelo estimado (I.3.3), por tanto, **subpredijo** el gasto de consumo real por casi 174 000 millones de dólares. Se diría que el **error de predicción** es de aproximadamente 174 000 millones de dólares, que representa alrededor de 1.5% del valor real del PIB para 2006. Cuando analicemos a profundidad el modelo de regresión lineal en los siguientes capítulos, trataremos de averiguar si un error de esa naturaleza es "pequeño" o "grande". Pero lo que ahora importa es observar que tales errores de predicción son inevitables, dada la naturaleza estadística del análisis.

Existe otro uso del modelo estimado (I.3.3). Suponga que el presidente decide proponer una reducción del impuesto sobre la renta. ¿Cuál será el efecto de dicha política en el ingreso y por consiguiente en el gasto de consumo, y a final de cuentas en el empleo?

Suponga que como resultado de estos cambios de política se incrementa el gasto en inversión. ¿Cuál será el efecto en la economía? De acuerdo con la teoría macroeconómica, el cambio en el ingreso generado por un cambio equivalente a un dólar, por ejemplo, en el gasto en inversión está dado por el **multiplicador del ingreso (*M*)**, el cual se define como

$$M = \frac{1}{1 - \text{PMC}} \tag{I.3.5}$$

Si utilizamos la PMC de 0.72 obtenida en la ecuación (I.3.3), este multiplicador se convierte en $M = 3.57$. Es decir, un aumento (o reducción) de un dólar en la inversión *al final* generará un incremento (o reducción) de más de tres veces en el ingreso; advierta que el multiplicador demora algún tiempo en actuar.

El valor crítico en este cálculo es la PMC, pues *M* depende de él. Y este valor estimado de la PMC se obtiene de modelos de regresión como el de la ecuación (I.3.3). Así, un valor estimado cuantitativo de la PMC proporciona información valiosa para fines de políticas públicas. Al conocer la PMC, se puede predecir el curso futuro del ingreso, el gasto de consumo y el empleo que sigue a un cambio en las políticas fiscales del gobierno.

[14] Había datos disponibles sobre el GCP y el PIB para 2006, pero los omitimos a propósito con el objeto de ilustrar el tema que estudiamos en esta sección. Como veremos en los capítulos subsiguientes, es buena idea guardar parte de los datos con el objeto de averiguar cómo predicen el modelo ajustado las observaciones ajenas a la muestra.

## 8. Uso del modelo para fines de control o de políticas

Suponga que tenemos la función keynesiana de consumo estimada dada en (I.3.3). Suponga además que el gobierno considera que un nivel de gasto de aproximadamente 8 750 (miles de millones de dólares de 2000) mantendrá la tasa de desempleo en su nivel actual de cerca de 4.2 por ciento (estimación para principios del 2006). ¿Qué nivel de ingreso garantizará la cantidad de gasto de consumo fijado como meta?

Si los resultados de la regresión dados en la ecuación (I.3.3) parecen razonables, la aritmética simple mostrará que

$$8\,750 = -299.5913 + 0.7218(PIB_{2006}) \qquad \text{(I.3.6)}$$

que da $X = 12\,537$, aproximadamente. Es decir, un nivel de ingresos de alrededor de 12 537 (miles de millones) de dólares, con una PMC de cerca de 0.72, producirá un gasto aproximado de 8 750 millones de dólares.

Como indican estos cálculos, un modelo estimado sirve para fines de control o de políticas públicas. Mediante una mezcla apropiada de política fiscal y monetaria, el gobierno puede manejar la **variable de control *X*** para producir el nivel deseado de la **variable objetivo *Y***.

La figura I.4 resume la anatomía de la creación de los modelos econométricos clásicos.

## Elección entre modelos rivales

Cuando una dependencia gubernamental (digamos, el Departamento de Comercio de Estados Unidos) recopila datos económicos, como los de la tabla I.1, no necesariamente tiene una teoría económica en mente. Por tanto, ¿cómo sabe en realidad que los datos respaldan la teoría keynesiana de consumo? ¿Se debe acaso a que la función consumo keynesiana (es decir, la línea de regresión) de la figura I.3 se aproxima mucho a los puntos reales que representan a los datos? ¿Será posible que otro modelo (teoría) de consumo se ajuste igual de bien a los datos? Por ejemplo,

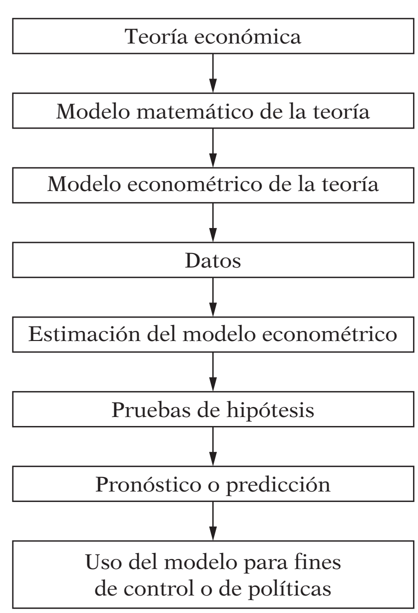



**FIGURA I.4**
Anatomía de la creación de modelos econométricos.

Milton Friedman elaboró un modelo de consumo, la *hipótesis de ingreso permanente*.[15] Robert Hall también creó un modelo de consumo, llamado *hipótesis del ciclo de vida del ingreso permanente*.[16] ¿Alguno o ambos modelos pueden también ajustarse a los datos de la tabla I.1?

En resumen, la interrogante con que se enfrenta en la práctica un investigador es: ¿cómo elegir entre modelos o hipótesis que compiten entre sí, dado un fenómeno determinado, como la relación entre consumo e ingreso? Como observa Miller:

> Ningún encuentro con los datos significa un paso adelante hacia la confirmación genuina, a menos que la hipótesis se las arregle mejor con esos datos que algún rival natural. . . . Lo que fortalece aquí a una hipótesis es una victoria que, al mismo tiempo, es una derrota para una posible rival.[17]

Entonces, ¿cómo elegir entre los varios modelos o hipótesis en disputa? Aquí Clive Granger da un consejo que vale la pena:[18]

> Me gustaría proponer que en el futuro, cuando a uno se le presente una nueva teoría o modelo empírico, se plantee las siguientes preguntas:
>
> *i*) ¿Qué propósito tiene? ¿Qué tipo de decisiones económicas ayuda a tomar?
>
> *ii*) ¿Existe alguna evidencia presente que me permita evaluar su calidad en comparación con teorías o modelos alternos?
>
> Pienso que si se les da la debida atención a estos planteamientos se fortalecerá la investigación y el análisis económicos.

Conforme avancemos en este libro, saldrán al paso diversas hipótesis que compiten entre sí y que tratan de explicar varios fenómenos económicos. Por ejemplo, los estudiantes de economía conocen ya el concepto de la función producción, que representa básicamente una relación entre la producción y los insumos (capital y trabajo). En la bibliografía, dos funciones producción muy conocidas son la de *Cobb-Douglas* y la de *elasticidad constante de sustitución*. Con los datos de producción e insumos tendremos que averiguar cuál de las dos funciones producción, si acaso alguna lo hace, se ajusta bien a los datos.

La metodología econométrica clásica, consistente en los ocho pasos que acabamos de presentar, es neutral en el sentido de que sirve para probar cualquiera de estas hipótesis rivales.

¿Es posible elaborar una metodología lo bastante amplia para abarcar hipótesis contendientes? La respuesta implica un tema polémico e intrincado que analizaremos en el capítulo 13, tras entender la teoría econométrica necesaria.

## I.4 Tipos de econometría

Como deja entrever el esquema de clasificación en la figura I.5, la econometría se divide en dos amplias categorías: **econometría teórica** y **econometría aplicada**. En cada categoría se puede tratar la materia según la tradición **clásica** o la **bayesiana**. En este libro destacamos el enfoque clásico. Para el enfoque bayesiano, el lector puede consultar las referencias al final del capítulo.

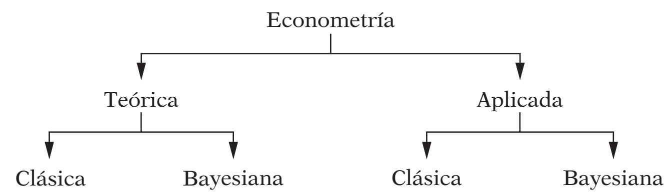
[15] Milton Friedman, *A Theory of Consumption Function*, Princeton University Press, Princeton, Nueva Jersey, 1957.

[16] R. Hall, "Stochastics Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy*, 1978, vol. 86, pp. 971-987.

[17] R.W. Miller, *Fact and Method: Explanation, Confirmation, and Reality in the Nature and Social Sciences*, Princeton University Press, Princeton, Nueva Jersey, 1978, p. 176.

[18] Clive W.J. Granger, E*mpirical Modeling in Economics*, Cambridge University Press, Gran Bretaña, 1999, p. 58.

**FIGURA I.5**
Categorías de la econometría.




La econometría teórica se relaciona con la elaboración de métodos apropiados para medir las relaciones económicas especificadas por los modelos econométricos. En este aspecto, la econometría se apoya en gran medida en la estadística matemática. Por ejemplo, un método muy popular en este libro es el de **mínimos cuadrados**. La econometría teórica debe expresar los supuestos de este método, sus propiedades y lo que les sucede cuando no se cumplen uno o más de los supuestos del método.

En la econometría aplicada utilizamos herramientas de la econometría teórica para estudiar algunos campos especiales de la economía y los negocios, como la función de producción, la función de inversión, las funciones de demanda y de oferta, la teoría de portafolio, etcétera.

Este libro se refiere en gran parte al desarrollo de los métodos econométricos, sus supuestos, usos y limitaciones. Ilustramos estos métodos con ejemplos en diversas áreas de la economía y los negocios. Pero éste no es un libro de econometría aplicada en el sentido de que investigue a fondo un campo particular de aplicación económica. Para esa labor existen textos especializados. Al final de esta obra proporcionamos referencias de algunos de ellos.

## I.5 Requisitos matemáticos y estadísticos

A pesar de que este libro está escrito en un nivel elemental, el autor supone que el lector conoce los conceptos básicos de la estimación estadística y las pruebas de hipótesis. Sin embargo, para quienes deseen refrescar sus conocimientos, en el **apéndice A** se ofrece una revisión amplia pero no técnica de los conceptos estadísticos básicos de esta obra. Respecto de las matemáticas, es deseable, aunque no esencial, estar más o menos al día con las nociones de cálculo diferencial. Si bien la mayoría de los textos universitarios de econometría emplea con libertad el álgebra matricial, deseo aclarar que este libro no la requiere. Sostengo la firme convicción de que las ideas fundamentales de econometría pueden transmitirse sin álgebra matricial. Sin embargo, para el beneficio del estudiante amigo de las matemáticas, el **apéndice C** resume la teoría de regresión básica en notación matricial. Para estos estudiantes, el **apéndice B** proporciona un resumen sucinto de los principales resultados del álgebra matricial.

## I.6 La función de la computadora

El análisis de regresión, herramienta de uso diario de la econometría, no sería posible hoy en día sin la computadora y el software estadístico. (Créanme, yo crecí en la generación de la regla de cálculo.) Por fortuna, ya existen muchos paquetes de regresión excelentes, tanto para las computadoras centrales (*mainframe*) como para las microcomputadoras, y con el tiempo la lista crece. Los paquetes de software de regresión, como **ET**, **LIMDEP**, **SHAZAM**, **MICRO TSP**, **MINITAB**, **EVIEWS, SAS**, **SPSS**, **BMD**, **STATA**, **Microfit** y **PcGive** tienen la mayoría de las técnicas econométricas y las pruebas analizadas en este libro.

En esta obra ocasionalmente pediremos al lector realizar experimentos **Monte Carlo** con uno o más paquetes estadísticos. Los experimentos Monte Carlo son ejercicios "divertidos" que capacitarán al lector para apreciar las propiedades de diversos métodos estadísticos analizados en este libro. Detallaremos sobre los experimentos Monte Carlo en las secciones pertinentes.

## I.7 Lecturas sugeridas

El tema de la metodología econométrica es vasto y controvertido. Para los interesados en este tema, sugiero los siguientes libros:

Neil de Marchi y Christopher Gilbert, eds., *History and Methodology of Econometrics*, Oxford University Press, Nueva York, 1989. En esta colección de lecturas se analizan los primeros trabajos sobre metodología econométrica. El análisis se extiende al método británico de la econometría relacionado con cifras de series de tiempo, es decir, datos recopilados a través de un periodo determinado.

Wojciech W. Charemza y Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, Edward Elgar, Hants, Inglaterra, 1997. Los autores critican el método tradicional de la econometría y dan una exposición detallada de nuevos enfoques a la metodología econométrica.

Adrian C. Darnell y J. Lynne Evans, *The Limits of Econometrics*, Edward Elgar, Hants, Inglaterra, 1990. Este libro presenta un análisis, en cierta medida equilibrado, de los diversos enfoques metodológicos a la econometría, con una renovada fidelidad a la metodología econométrica tradicional.

Mary S. Morgan, *The History of Econometric Ideas*, Cambridge University Press, Nueva York, 1990. La autora proporciona una perspectiva histórica excelente sobre la teoría y la práctica de la econometría, con un análisis a fondo de las primeras contribuciones de Haavelmo (Premio Nobel de Economía 1990) a la econometría. Con el mismo espíritu, David F. Hendry y Mary S. Morgan antologaron escritos seminales para la econometría en *The Foundation of Econometric Analisis*, Cambridge University Press, Gran Bretaña, 1995, con el objeto de mostrar la evolución de las ideas econométricas a través del tiempo.

David Colander y Reuven Brenner, eds., *Educating Economists*, University of Michigan Press, Ann Arbor, Michigan, 1992. El texto presenta un punto de vista crítico, en ocasiones agnóstico, de la enseñanza y práctica de la economía.

Para consultar sobre los temas de estadística y econometría bayesianas, los siguientes libros pueden ser útiles: John H. Dey, *Data in Doubt*, Basil Blackwell, Oxford, University Press, Inglaterra, 1985; Peter M. Lee, *Bayesian Statistics: An Introduction*, Oxford University Press, Inglaterra, 1989; y Dale J. Porier, *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge, Massachusetts, 1995. Una referencia avanzada es Arnold Zellner, *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Nueva York, 1971. Otro libro de consulta avanzada es *Palgrave Handbook of Econometrics.* Volumen I. *Econometric Theory,* Terence C. Mills y Kerry Patterson, eds., Palgrave Macmillan, Nueva York, 2007.

Parte 1

# Modelos de regresión uniecuacionales

En la parte 1 de este texto se presentan los modelos de regresión uniecuacionales. En estos modelos se expresa una variable, llamada *dependiente*, como función lineal de una o más variables, llamadas *explicativas*. En modelos de este tipo se supone que si existen relaciones causales entre las variables dependientes y las explicativas, éstas van en una sola dirección: de las variables explicativas a la variable dependiente.

En el capítulo 1 se hace una exposición relacionada con la interpretación, tanto histórica como moderna, del término *regresión* y se ilustran las diferencias entre las dos interpretaciones con diversos ejemplos tomados de la economía y de otros campos.

En el capítulo 2 se presentan algunos conceptos fundamentales del análisis de regresión con ayuda del modelo de regresión lineal con dos variables, en el cual la variable dependiente se expresa como función lineal de una sola variable explicativa.

En el capítulo 3 continúa el manejo del modelo con dos variables y se introduce lo que se conoce como el *modelo clásico de regresión lineal*, que tiene diversos supuestos simplificadores. Con estos supuestos se presenta el método de *mínimos cuadrados ordinarios* (MCO) para estimar los parámetros del modelo de regresión con dos variables. La aplicación del método de MCO es sencilla y tiene algunas propiedades estadísticas muy convenientes.

En el capítulo 4 se introduce el modelo clásico de regresión lineal *normal* (de dos variables), modelo que supone que la variable aleatoria dependiente sigue una distribución de probabilidad normal. Con este supuesto los estimadores MCO obtenidos en el capítulo 3 adquieren algunas propiedades estadísticas más sólidas que las de los modelos clásicos de regresión lineal no normales. Estas propiedades permiten la inferencia estadística y, en particular, las pruebas de hipótesis.

El capítulo 5 se dedica a las pruebas de hipótesis, y se pretende averiguar si los coeficientes de regresión estimados son compatibles con los valores hipotéticos de tales coeficientes, valores hipotéticos sugeridos por la teoría y/o por el trabajo empírico previo.

En el capítulo 6 se consideran algunas extensiones del modelo de regresión con dos variables. En particular, se analizan temas como: 1) regresión a través del origen, 2) escalas y unidades de medición, y 3) formas funcionales de modelos de regresión, como doblelogarítmicos, semilogarítmicos y recíprocos.

En el capítulo 7 se considera el modelo de regresión múltiple, en el cual hay más de una variable explicativa, y se muestra cómo se extiende el método MCO para estimar los parámetros de tales modelos.

En el capítulo 8 se amplían los conceptos del capítulo 5 al modelo de regresión múltiple y se señalan algunas complicaciones propias de diversas variables explicativas.

El capítulo 9, que trata sobre variables explicativas dicótomas o cualitativas, concluye la primera parte del texto. Este capítulo destaca que no todas las variables explicativas necesitan ser cuantitativas (por ejemplo, en escala de razón). Variables como género, raza, religión, nacionalidad y lugar de residencia no son cuantificables de manera directa, si bien desempeñan un valioso papel en la explicación de muchos fenómenos económicos.

# Capítulo 1

# Naturaleza del análisis de regresión

Como se mencionó en la introducción, la regresión es una herramienta fundamental de la econometría; en este capítulo se considera muy brevemente la naturaleza de este instrumento.

## 1.1 Origen histórico del término *regresión*

Francis Galton acuñó el término *regresión*. En un famoso ensayo, Galton planteó que, a pesar de la tendencia de los padres de estatura alta a procrear hijos altos y los padres de estatura baja, hijos bajos, la estatura promedio de los niños de padres de una estatura determinada tendía a desplazarse, o "regresar", a la estatura promedio de la población total.[1] En otras palabras, la estatura de los hijos de padres inusualmente altos o inusualmente bajos tiende a dirigirse a la estatura promedio de la población. La *ley de regresión universal* de Galton fue confirmada por su amigo Karl Pearson, quien reunió más de mil registros de estaturas de miembros de grupos familiares.[2] Pearson descubrió que la estatura promedio de los hijos de un grupo de padres de estatura alta era menor que la estatura de sus padres, y que la estatura promedio de los hijos de un grupo de padres de estatura baja era mayor que la estatura de sus padres; es decir, se trata de un fenómeno mediante el cual los hijos altos e hijos bajos "regresan" por igual a la estatura promedio de todos los demás. En palabras de Galton, se trata de una "regresión a la mediocridad".

## 1.2 Interpretación moderna de la regresión

La interpretación moderna de la regresión es, sin embargo, muy diferente. En términos generales, se afirma que:

> El análisis de regresión trata del estudio de la dependencia de una variable (*variable dependiente*) respecto de una o más variables (*variables explicativas*) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las segundas.

[1] Francis Galton, "Family Likeness in Stature", *Proceedings of Royal Society*, Londres, vol. 40, 1886, pp. 42-72.

[2] K. Pearson y A. Lee, "On the Laws of Inheritance", *Biometrika*, vol. 2, noviembre de 1903, pp. 357-462.

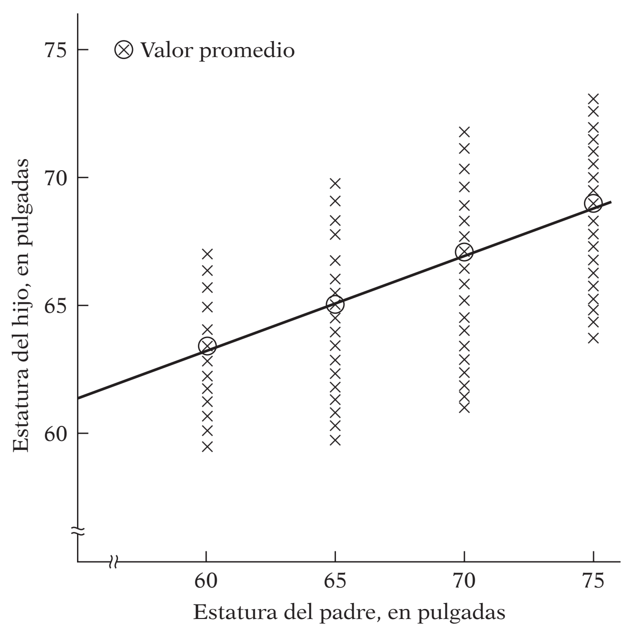
La importancia trascendental de este enfoque del análisis de regresión se verá claramente sobre la marcha, pero algunos ejemplos sencillos aclararán este concepto básico.

## Ejemplos

1. Considere de nuevo la ley de regresión universal de Galton. A él le interesaba averiguar las razones de la estabilidad en la distribución de estaturas dentro de una población. En el enfoque moderno, la preocupación no es esta explicación, sino averiguar cómo cambia la estatura *promedio* de los hijos dada la estatura de los padres. En otras palabras, lo que interesa es predecir la estatura promedio de los hijos a partir de la estatura de sus padres. Para ver cómo hacerlo, considere la figura 1.1, que corresponde a un **diagrama de dispersión**. La figura muestra la distribución de las estaturas de los hijos en una población hipotética, correspondiente al conjunto de valores dados o *fijos* de las estaturas de los padres. Observe que, para cualquier estatura de un padre, existe un *rango* (distribución) de estaturas de los hijos. Sin embargo, observe también que, a pesar de la variabilidad de la estatura de los hijos conforme al valor de la estatura de los padres, la estatura promedio de los hijos aumenta, por lo general, en la medida en que lo hace la estatura de los padres. Para demostrar esto con claridad, las cruces dentro de los círculos en la figura indican la estatura *promedio* de los hijos que corresponde a una estatura determinada de los padres. Estos promedios se conectan para obtener la línea recta de la figura. Esta línea, como veremos, se conoce como **recta de regresión**. Dicha recta muestra que el *promedio* de la estatura de los hijos aumenta conforme crece la de los padres.[3]

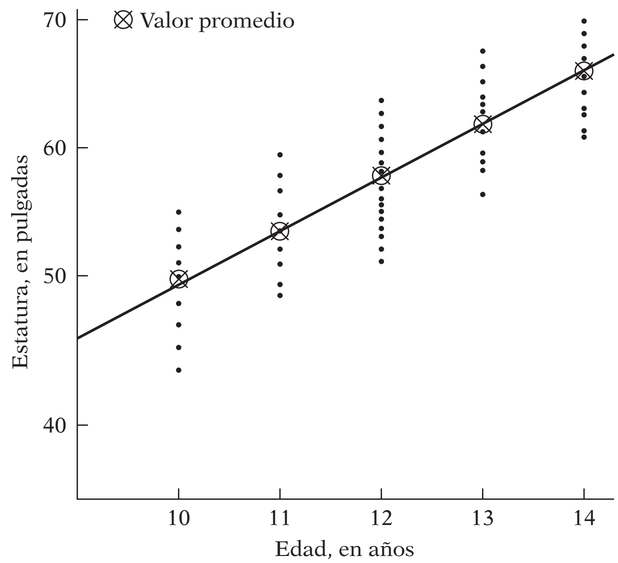
2. Considere el diagrama de dispersión en la figura 1.2, que presenta la distribución de una población hipotética de estaturas de niños en edades *fijas*. Observe que existe un rango (distribución) de estaturas correspondiente a cada edad. Es obvia la improbabilidad de que todos los niños de una edad determinada tengan estaturas idénticas. Pero, *en promedio*, la estatura se incrementa con la edad (por supuesto, hasta cierta edad), que se ve con claridad al trazar una recta (la recta de

**FIGURA 1.1**
Distribución hipotética de las estaturas de los hijos correspondientes a las estaturas de los padres.




[3] En esta etapa de estudio del tema, denominaremos a esta recta de regresión simplemente *recta que conecta el valor de la media, o promedio, de la variable dependiente (la estatura de los hijos) que corresponde a un valor dado de la variable explicativa (la estatura de los padres).* Observe que esta recta tiene una pendiente positiva; pero la pendiente es menor que 1, lo cual está de acuerdo con el concepto de Galton de regresión a la mediocridad. (¿Por qué?)

**FIGURA 1.2**
Distribución hipotética de estaturas correspondientes a edades seleccionadas.




regresión) por los puntos dentro de los círculos, los cuales representan la estatura promedio de determinadas edades. Por consiguiente, si se conoce la edad, se predice la estatura promedio de dicha edad mediante la recta de regresión.

3. Al considerar lo referente a la economía, a un economista quizá le interese estudiar la dependencia del consumo personal respecto del ingreso personal neto disponible (después de impuestos). Con un análisis de este tipo se calcula la propensión marginal a consumir (PMC), es decir, el cambio promedio del consumo ante un cambio, digamos, de un dólar en el ingreso real (ver la figura 1.3).

4. Un monopolista que puede fijar el precio o la producción (pero no ambos factores) tal vez desee conocer la demanda de un producto con diversos precios. Tal experimento permite estimar la **elasticidad del precio** (es decir, la respuesta a variaciones del precio) de la demanda del producto y permite determinar el precio que maximiza las ganancias.

5. Un economista laboral quizá desee estudiar la tasa de cambio de los salarios monetarios o nominales en relación con la tasa de desempleo. Las cifras históricas aparecen en el diagrama de dispersión de la figura 1.3. La curva de esta figura es un ejemplo de la célebre *curva de Phillips*, que relaciona los cambios en los salarios nominales con la tasa de desempleo. Un diagrama de dispersión de este tipo permite al economista laboral predecir el cambio promedio en los salarios nominales con una cierta tasa de desempleo. Tal conocimiento sirve para establecer supuestos sobre el proceso inflacionario en una economía, pues es probable que los incrementos en los salarios monetarios se reflejen en incrementos de precios.

6. En la economía monetaria se sabe que, si se mantienen constantes otros factores, cuanto mayor sea la tasa de inflación $\pi$, menor será la proporción $k$ del ingreso que la gente deseará mantener en forma de dinero, como se deduce de la figura 1.4. La pendiente de esta recta representa el cambio en $k$ con un cambio en la tasa de inflación. Un análisis cuantitativo de esta relación permite al economista predecir la cantidad de dinero, como proporción del ingreso, que la gente deseará mantener con diversas tasas de inflación.

7. El director de marketing de una compañía tal vez quiera conocer la relación entre la demanda del producto de su compañía con el gasto de publicidad, por ejemplo. Un estudio de este tipo es de gran ayuda para encontrar la **elasticidad de la demanda** respecto de los gastos publicitarios, es decir, el cambio porcentual de la demanda en respuesta a un cambio de 1 por ciento, por ejemplo, en el presupuesto de publicidad. Saber esto sirve para determinar el presupuesto "óptimo" de publicidad.

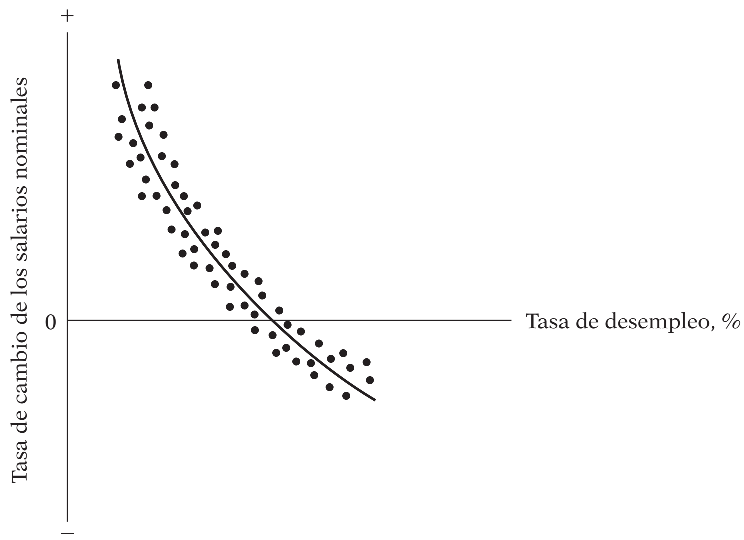



**FIGURA 1.3**
Curva hipotética de Phillips.

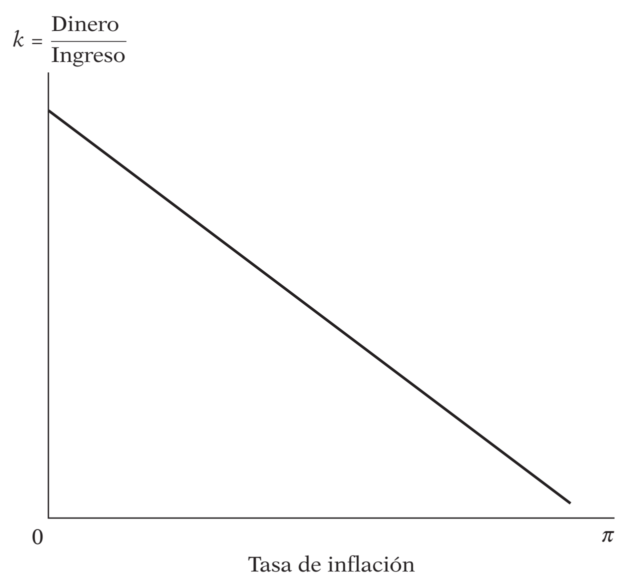



**FIGURA 1.4**
Tenencia de dinero en relación con la tasa de inflación $\pi$.

8. Por último, un agrónomo tal vez se interese en estudiar la relación entre el rendimiento de un cultivo, digamos de trigo, y la temperatura, lluvia, cantidad de sol y fertilizantes. Un análisis de dependencia de ese tipo facilitaría la predicción o el pronóstico del rendimiento medio del cultivo según la información sobre las variables explicativas.

El lector puede proporcionar una amplia gama de ejemplos similares de la dependencia de una variable respecto de otra o más variables. Las técnicas del análisis de regresión que se explican en este texto están diseñadas especialmente para estudiar dicha dependencia entre variables.

## 1.3 Relaciones estadísticas y relaciones deterministas

En los ejemplos de la sección 1.2 se observa que en el análisis de regresión interesa lo que se conoce como dependencia *estadística* entre variables, no así la *funcional* o *determinista*, propia de la física clásica. En las relaciones estadísticas entre variables se analizan, en esencia, variables **aleatorias** o **estocásticas**,[4] es decir, variables con distribuciones de probabilidad. Por otra parte, en la dependencia funcional o determinista también se manejan variables, pero no son aleatorias o estocásticas.

Por ejemplo, el rendimiento de un cultivo depende de la temperatura, lluvia, Sol y fertilizantes, y dicha dependencia es de naturaleza estadística porque las variables explicativas, si bien son importantes, no permiten al agrónomo predecir en forma exacta el rendimiento del cultivo debido a los errores propios de la medición de estas variables y a otra serie de factores (variables) que en conjunto afectan el rendimiento pero son difíciles de identificar individualmente. De esta manera, habrá alguna variabilidad "intrínseca" o aleatoria en la variable dependiente, el rendimiento del cultivo, que no puede explicarse en su totalidad sin importar cuántas variables explicativas se consideren.

Los fenómenos deterministas, por otra parte, implican relaciones como la ley de la gravedad de Newton, la cual establece que toda partícula en el universo atrae a cualquier otra partícula con una fuerza directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia entre ellas. En términos matemáticos, $F = k(m_1 m_2/r^2)$, donde $F =$ fuerza, $m_1$ y $m_2$ son las masas de las dos partículas, $r =$ distancia y $k =$ constante de proporcionalidad. Otro ejemplo es el de la ley de Ohm, la cual postula que para conductores metálicos dentro de un intervalo limitado de temperatura, la corriente $C$ es proporcional al voltaje $V$; es decir, $C = (\frac{1}{k})V$, donde $\frac{1}{k}$ es la constante de proporcionalidad. Otros ejemplos de relaciones deterministas son la ley de los gases de Boyle, la ley de la electricidad de Kirchhoff y la ley del movimiento de Newton.

En este texto no interesan tales relaciones deterministas. Por supuesto, de haber errores de medición, por ejemplo, en la $k$ de la ley de la gravedad de Newton, la relación que de otra forma habría sido determinista se convierte en una relación estadística. En esta situación es posible predecir la fuerza en forma aproximada sólo a partir de un valor dado de $k$ (y $m_1$, $m_2$ y $r$), el cual contiene errores. La variable $F$ se convierte en este caso en aleatoria.

## 1.4 Regresión y causalidad

A pesar de que el análisis de regresión tiene que ver con la dependencia de una variable respecto de otras variables, esto no implica causalidad necesariamente. En palabras de Kendall y Stuart: "Una relación estadística, por más fuerte y sugerente que sea, nunca podrá establecer una conexión causal: nuestras ideas de causalidad deben provenir de estadísticas externas y, en último término, de una u otra teoría".[5]

[4] La palabra *estocástico* viene de la voz griega *stokhos,* que significa "centro del blanco". El resultado de lanzar dardos sobre un tablero es un proceso estocástico, es decir, un proceso lleno de tiros fallidos.

[5] M. G. Kendall y A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin Publishers, Nueva York, 1961, vol. 2, cap. 26, p. 279.

En el ejemplo del rendimiento del cultivo citado, no hay una *razón estadística* para suponer que la lluvia no depende del rendimiento del cultivo. Considerar que el rendimiento del cultivo depende de la lluvia (entre otras cosas) se debe a cuestiones no estadísticas: el sentido común indica que la relación no puede ser a la inversa, pues no es posible controlar la lluvia mediante el rendimiento del cultivo.

En todos los ejemplos de la sección 1.2, lo que se debe notar es que **una relación estadística por sí misma no puede, por lógica, implicar causalidad**. Para aducir causalidad se debe acudir a consideraciones *a priori* o teóricas. Así, en el tercer ejemplo, es posible recurrir a la teoría económica para afirmar que el consumo depende del ingreso real.[6]

## 1.5 Regresión y correlación

El **análisis de correlación** se relaciona de manera estrecha con el de regresión, aunque conceptualmente los dos son muy diferentes. En el análisis de correlación, el objetivo principal es medir la *fuerza* o el *grado* de *asociación lineal* entre dos variables. El **coeficiente de correlación**, que veremos en detalle en el capítulo 3, mide esta fuerza de asociación (lineal): por ejemplo, si se desea encontrar la correlación (coeficiente) entre el hábito de fumar y el cáncer del pulmón; entre las calificaciones en exámenes de estadística y en exámenes de matemáticas; entre las calificaciones de bachillerato y de la universidad, y así sucesivamente. En el análisis de regresión, como ya mencionamos, no interesa ese tipo de medición. En cambio, se trata de estimar o predecir el valor promedio de una variable con base en los valores fijos de otras. Así, quizá se desee predecir el promedio de las calificaciones en un examen de estadística a partir de la calificación de un estudiante en un examen de matemáticas.

La regresión y la correlación presentan diferencias fundamentales que vale la pena mencionar. En el análisis de regresión hay una asimetría en el tratamiento a las variables dependientes y explicativas. Se supone que la variable dependiente es estadística, aleatoria o estocástica, es decir, que tiene una distribución de probabilidad. Por otra parte, se asume que las variables explicativas tienen valores fijos (en muestras repetidas),[7] lo cual es explícito en la definición de regresión de la sección 1.2. Así, en la figura 1.2 se supuso que la variable de edad era fija en los niveles dados y se obtuvieron medidas de estatura en esos niveles. En el análisis de correlación, por otra parte, se tratan dos variables cualesquiera en forma simétrica; no hay distinción entre las variables dependiente y explicativa. Después de todo, la correlación entre las calificaciones de los exámenes de matemáticas y de estadística es la misma que la existente entre calificaciones de exámenes de estadística y de matemáticas. Además, las dos variables se consideran aleatorias. Como veremos, la mayor parte de la teoría de correlación parte del supuesto de aleatoriedad de las variables, mientras que la mayor parte de la teoría de regresión que expondremos en este texto está condicionada al supuesto de que la variable dependiente es estocástica y que las variables explicativas son fijas o no estocásticas.[8]

[6] Pero, como veremos en el capítulo 3, el análisis clásico de regresión se basa en el supuesto de que el modelo del análisis es el correcto. Por consiguiente, la dirección de la causalidad puede estar implícita en el modelo postulado.

[7] Es de crucial importancia notar que las variables explicativas pueden ser intrínsecamente estocásticas, pero, para fines del análisis de regresión, suponemos que sus valores son fijos en el muestreo repetido (es decir, que *X* toma los mismos valores en diversas muestras), de modo que, en efecto, no resultan aleatorias ni estocásticas. Hay más sobre este tema en la sección 3.2 del capítulo 3.

[8] En el tratamiento avanzado de econometría se puede ser más flexible con el supuesto de que las variables explicativas son no estocásticas (ver la introducción de la segunda parte).

## 1.6 Terminología y notación

Antes de proceder al análisis formal de la teoría de regresión abordaremos brevemente la terminología y la notación. En las publicaciones especializadas, los términos *variable dependiente* y *variable explicativa* se definen de varias maneras; a continuación se presenta una lista representativa:

| **Variable dependiente** | **Variable explicativa** |
|---|---|
| ⇕ | ⇕ |
| Variable explicada | Variable independiente |
| ⇕ | ⇕ |
| Predicha | Predictora |
| ⇕ | ⇕ |
| **Regresada** | **Regresora** |
| ⇕ | ⇕ |
| Respuesta | Estímulo |
| ⇕ | ⇕ |
| Endógena | Exógena |
| ⇕ | ⇕ |
| Resultado | Covariante |
| ⇕ | ⇕ |
| Variable controlada | Variable de control |

Aunque es cuestión de preferencia personal y tradición, en este texto se utiliza la terminología de variable dependiente/variable explicativa, o la más neutral de regresada y regresora.

Si se estudia la dependencia de una variable respecto de una única variable explicativa, como el consumo que depende del ingreso real, dicho estudio se conoce como **análisis de regresión** *simple*, o **con dos variables**. Sin embargo, si se estudia la dependencia de una variable respecto de más de una variable explicativa, como el rendimiento de un cultivo, la lluvia, la temperatura, el Sol y los fertilizantes, se trata de un **análisis de regresión múltiple**. En otras palabras, en una regresión de dos variables sólo hay una variable explicativa, mientras que en la regresión múltiple hay más de una variable explicativa.

El término **aleatorio** es sinónimo de **estocástico**. Como ya vimos, una variable aleatoria o estocástica es la que toma cualquier conjunto de valores, positivos o negativos, con una probabilidad dada.[9]

A menos que se indique lo contrario, la letra $Y$ representa la variable dependiente, y las $X$ ($X_1$, $X_2, \ldots, X_k$), las variables explicativas, con $X_k$ como la $k$-ésima variable explicativa. Los subíndices $i$ o $t$ denotan la observación o valor $i$-ésimo o $t$-ésimo. $X_{ki}$ (o $X_{kt}$) denota la $i$-ésima (o la $t$-ésima) observación de la variable $X_k$. $N$ (o $T$) representa el número total de observaciones o valores en la población, y $n$ (o $t$), el número total de observaciones en una muestra. Por convención, se utiliza el subíndice de observación $i$ para los **datos transversales** (es decir, información recopilada en un momento determinado), y el subíndice $t$, para **datos de series de tiempo** (es decir, información reunida a lo largo de un periodo). La naturaleza de datos transversales y de series de tiempo, así como el importante tema de la naturaleza y las fuentes de datos para el análisis empírico, se estudian en la siguiente sección.

[9] Hay una definición formal y más detalles en el **apéndice A**.

## 1.7 Naturaleza y fuentes de datos para el análisis económico[10]

El éxito de todo análisis econométrico depende a final de cuentas de la disponibilidad de los datos recopilados. Por consiguiente, es muy importante dedicar algún tiempo a estudiar la naturaleza, las fuentes y las limitaciones de los datos para el análisis empírico.

### Tipos de datos

Hay tres tipos de datos disponibles para el análisis empírico: **series de tiempo**, **series transversales** e **información combinada** (combinación de series de tiempo y transversales).

#### *Datos de series de tiempo*

Los datos de la tabla 1.1 son un ejemplo de datos de series de tiempo. Una *serie de tiempo* es un conjunto de observaciones sobre los valores de una variable en diferentes momentos. Tal información debe recopilarse en intervalos regulares, es decir, en forma **diaria** (precios de acciones, informes del tiempo, etc.), **semanal** (como cifras de oferta monetaria), **mensual** (tasa de desempleo, Índice de Precios al Consumidor [IPC], etc.), **trimestral** (como el PIB), **anual** (como los presupuestos del gobierno), **quinquenal** (como el censo de la industria manufacturera), o **decenal** (como los censos de población). Algunas veces los datos están disponibles por trimestre y por año, como los datos del PIB y del consumo. Con las computadoras de alta velocidad, ahora se recopilan datos en intervalos muy breves, por ejemplo, precios de acciones, que se obtienen literalmente de manera continua (o *cotización en tiempo real*).

Si bien los datos de series de tiempo se utilizan mucho en estudios econométricos, presentan algunos problemas especiales para los econometristas. Como veremos en los capítulos sobre **econometría de series de tiempo**, la mayor parte del trabajo empírico con datos de series de tiempo supone que éstas son **estacionarias**. Aunque es muy pronto para introducir el significado técnico preciso de estacionariedad, *en términos generales, una serie de tiempo es estacionaria si su media y varianza no varían sistemáticamente con el tiempo*. Para entender esto, observe, en la figura 1.5, el comportamiento de la oferta de dinero M1 en Estados Unidos durante el periodo del primero de enero de 1959 a septiembre de 1999. (Los datos reales se proporcionan en el ejercicio 1.4.) Como se observa, la oferta de dinero M1 presenta una **tendencia** ascendente constante, así como variabilidad con el transcurso de los años, lo cual indica que la serie de tiempo M1 no es estacionaria.[11] En el capítulo 21 se analiza a fondo este tema.

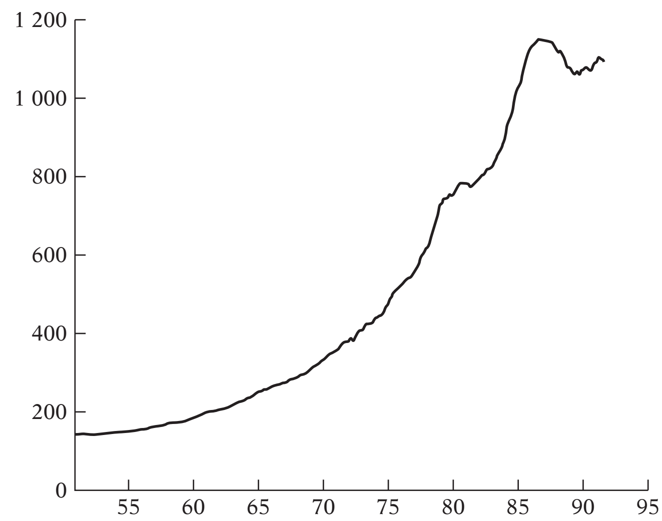
#### *Datos transversales*

Los datos transversales consisten en datos de una o más variables recopilados *en el mismo punto del tiempo*, como el censo de población realizado por la Oficina del Censo de Estados Unidos cada 10 años (el último fue en 2000), las encuestas de gastos del consumidor levantadas por la Universidad de Michigan y, sin duda, las encuestas de opinión de Gallup y diversas empresas especializadas. Un ejemplo concreto de datos transversales se presenta en la tabla 1.1, con datos sobre la producción y precios del huevo en Estados Unidos para los 50 estados durante 1990 y

[10] En Michael D. Intriligator, *Econometric Models, Techniques, and Applications*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1978, cap. 3, hay una explicación ilustrativa.

[11] Para que resulte más claro, los datos se dividen en cuatro periodos: enero de 1951 a diciembre de 1962; enero de 1963 a diciembre de 1974; enero de 1975 a diciembre de 1986; y enero de 1987 a septiembre de 1999. En estos subperiodos, los valores de la media para la oferta de dinero (con sus correspondientes desviaciones estándar, entre paréntesis) fueron, respectivamente: 165.88 (23.27), 323.20 (72.66), 788.12 (195.43) y 1 099 (27.84); todas las cifras representan miles de millones de dólares. Ésta es una indicación aproximada de que la oferta de dinero no fue estacionaria a lo largo de todo el periodo.

**FIGURA 1.5**
Oferta de dinero M1: Estados Unidos, enero de 1951-septiembre de 1999.




1991. Para cada año, los datos sobre los 50 estados son transversales. Así, en la tabla 1.1 aparecen dos muestras de corte transversal.

Así como los datos de series de tiempo crean problemas especiales (por la estacionariedad), los datos transversales también tienen sus propios problemas, en concreto, el de la *heterogeneidad.* En los datos de la tabla 1.1 se observa que hay algunos estados que producen grandes cantidades de huevo (como Pensilvania) y otros que producen muy poco (por ejemplo, Alaska). Cuando se incluyen unidades heterogéneas en un análisis estadístico, debe tenerse presente el **efecto de tamaño** o **de escala** con el fin de no mezclar manzanas con naranjas. Para ver esto con claridad, en la figura 1.6 se representan gráficamente los datos sobre la producción y los precios del huevo en los 50 estados de Estados Unidos en 1990. Esta figura muestra la amplia dispersión de las observaciones. En el capítulo 11 veremos que el efecto de escala puede ser importante al evaluar las relaciones entre variables económicas.

### *Datos combinados*

Los datos combinados reúnen elementos de series de tiempo y transversales. Los datos de la tabla 1.1 son datos combinados. Hay 50 observaciones transversales por año, y dos observaciones de series de tiempo sobre precios y producción de huevo por estado: un total de 100 observaciones combinadas. De igual forma, los datos del ejercicio 1.1 son combinados, pues el índice de precios al consumidor de cada país de 1980 a 2005 representa datos de series de tiempo, en tanto que los datos del IPC de los siete países correspondientes a un solo año son transversales. Los datos combinados consisten en 182 observaciones: 26 observaciones anuales para cada uno de los siete países.

### *Datos en panel, longitudinales o en micropanel*

Hay un tipo especial de datos combinados en el cual se estudia a través del tiempo la *misma* unidad transversal (por ejemplo, una familia o una empresa). Por ejemplo, el Departamento de Comercio de Estados Unidos realiza un censo de vivienda en intervalos periódicos. En cada encuesta periódica se entrevista a la misma unidad familiar (o a la gente que vive en la misma dirección) para averiguar si ha habido algún cambio en las condiciones de vivienda o financieras de esa unidad familiar desde la última encuesta. Los datos en panel que se obtienen de las entrevistas periódicas de la misma unidad familiar proporcionan información muy útil sobre la dinámica del comportamiento de las unidades familiares, como veremos en el capítulo 16.

**FIGURA 1.6**
Relación entre la producción y los precios del huevo, 1990.

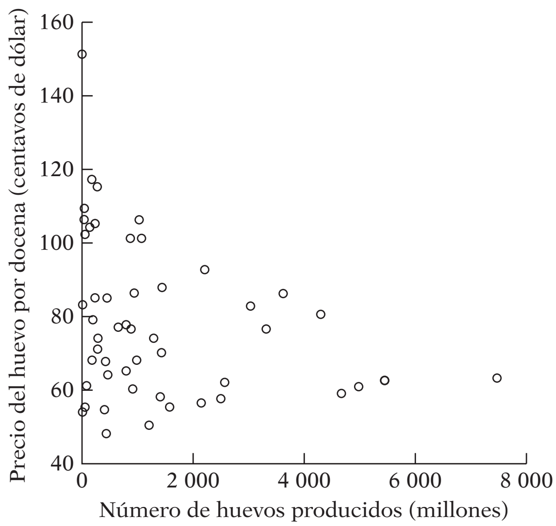



**TABLA 1.1** **Producción de huevo en Estados Unidos**

| Estado | $Y_1$ | $Y_2$ | $X_1$ | $X_2$ | Estado | $Y_1$ | $Y_2$ | $X_1$ | $X_2$ |
|---|---|---|---|---|---|---|---|---|---|
| AL | 2 206 | 2 186 | 92.7 | 91.4 | MT | 172 | 164 | 68.0 | 66.0 |
| AK | 0.7 | 0.7 | 151.0 | 149.0 | NE | 1 202 | 1 400 | 50.3 | 48.9 |
| AZ | 73 | 74 | 61.0 | 56.0 | NV | 2.2 | 1.8 | 53.9 | 52.7 |
| AR | 3 620 | 3 737 | 86.3 | 91.8 | NH | 43 | 49 | 109.0 | 104.0 |
| CA | 7 472 | 7 444 | 63.4 | 58.4 | NJ | 442 | 491 | 85.0 | 83.0 |
| CO | 788 | 873 | 77.8 | 73.0 | NM | 283 | 302 | 74.0 | 70.0 |
| CT | 1 029 | 948 | 106.0 | 104.0 | NY | 975 | 987 | 68.1 | 64.0 |
| DE | 168 | 164 | 117.0 | 113.0 | NC | 3 033 | 3 045 | 82.8 | 78.7 |
| FL | 2 586 | 2 537 | 62.0 | 57.2 | ND | 51 | 45 | 55.2 | 48.0 |
| GA | 4 302 | 4 301 | 80.6 | 80.8 | OH | 4 667 | 4 637 | 59.1 | 54.7 |
| HI | 227.5 | 224.5 | 85.0 | 85.5 | OK | 869 | 830 | 101.0 | 100.0 |
| ID | 187 | 203 | 79.1 | 72.9 | OR | 652 | 686 | 77.0 | 74.6 |
| IL | 793 | 809 | 65.0 | 70.5 | PA | 4 976 | 5 130 | 61.0 | 52.0 |
| IN | 5 445 | 5 290 | 62.7 | 60.1 | RI | 53 | 50 | 102.0 | 99.0 |
| IA | 2 151 | 2 247 | 56.5 | 53.0 | SC | 1 422 | 1 420 | 70.1 | 65.9 |
| KS | 404 | 389 | 54.5 | 47.8 | SD | 435 | 602 | 48.0 | 45.8 |
| KY | 412 | 483 | 67.7 | 73.5 | TN | 277 | 279 | 71.0 | 80.7 |
| LA | 273 | 254 | 115.0 | 115.0 | TX | 3 317 | 3 356 | 76.7 | 72.6 |
| ME | 1 069 | 1 070 | 101.0 | 97.0 | UT | 456 | 486 | 64.0 | 59.0 |
| MD | 885 | 898 | 76.6 | 75.4 | VT | 31 | 30 | 106.0 | 102.0 |
| MA | 235 | 237 | 105.0 | 102.0 | VA | 943 | 988 | 86.3 | 81.2 |
| MI | 1 406 | 1 396 | 58.0 | 53.8 | WA | 1 287 | 1 313 | 74.1 | 71.5 |
| MN | 2 499 | 2 697 | 57.7 | 54.0 | WV | 136 | 174 | 104.0 | 109.0 |
| MS | 1 434 | 1 468 | 87.8 | 86.7 | WI | 910 | 873 | 60.1 | 54.0 |
| MO | 1 580 | 1 622 | 55.4 | 51.5 | WY | 1.7 | 1.7 | 83.0 | 83.0 |

*Notas:* $Y_1$ = huevos producidos en 1990 (millones).
$Y_2$ = huevos producidos en 1991 (millones).
$X_1$ = precio por docena (centavos de dólar) en 1990.
$X_2$ = precio por docena (centavos de dólar) en 1991.
Fuente: *World Almanac*, 1993, p. 119. Los datos provienen del Economic Research Service, U.S. Department of Agriculture.

Como ejemplo concreto considere los datos proporcionados en la tabla 1.2. Los datos de la tabla, recopilados originalmente por Y. Grunfeld, se refieren a la inversión real, al valor real de la empresa y al capital accionario real de cuatro empresas estadounidenses, a saber, General Electric (GM), U.S. Steel (US), General Motors (GM) y Westinghouse (WEST), de 1935 a 1954.[12] En virtud de que los datos corresponden a varias empresas y se recopilaron a lo largo de varios años, se trata de un ejemplo clásico de datos en panel. En esta tabla, el número de observaciones de cada empresa es el mismo, pero no siempre ocurre así. Si todas las empresas tienen el mismo número de observaciones, se tiene lo que se conoce como **panel balanceado**. Si el número de observaciones no es igual para cada compañía, se llama **panel desbalanceado**. En el capítulo 16, Modelos de regresión con datos en panel, examinaremos estos datos y cómo estimar estos modelos.

El propósito de Grunfeld cuando recopiló estos datos fue investigar cómo depende la inversión bruta real ($I$) del valor real de la empresa ($F$) un año antes y del capital accionario real ($C$) un año antes. Como las compañías de esta muestra operan en el mismo mercado de capital, Grunfeld las estudió en conjunto para averiguar si tenían funciones de inversión parecidas.

## Fuentes de datos[13]

Los datos para el análisis empírico pueden provenir de una dependencia gubernamental (por ejemplo, el Departamento de Comercio), un organismo internacional (el Fondo Monetario Internacional [FMI] o el Banco Mundial), una organización privada (por ejemplo, Standard & Poor's) o un particular. Hay miles de agencias de este tipo que recopilan datos para uno u otro fin.

### *Internet*

Internet revolucionó la labor de recopilación de datos. Si uno "navega" por la red en los motores de búsqueda con sólo una palabra o frase (por ejemplo, tipos de cambio), se verá inundado con todo tipo de fuentes de datos. En el **apéndice E** se mencionan algunos sitios Web que suministran todo tipo de información financiera y económica, y que se visitan con mayor frecuencia. La mayoría de los datos se descarga con un costo mínimo. Conviene incluir en la lista de Favoritos, los sitios Web que brinden datos económicos útiles.

Los datos recopilados por estas organizaciones pueden ser de naturaleza **experimental** o **no experimental**. En los datos experimentales, frecuentes en las ciencias naturales, el investigador suele recabar los datos con algunos factores constantes, con el fin de evaluar el efecto de otros en un fenómeno dado. Por ejemplo, al estimar el efecto de la obesidad en la presión arterial, el investigador recopilaría los datos y mantendría constantes los hábitos de las personas respecto de comer, fumar y beber para reducir la influencia de estas variables en la presión arterial.

En las ciencias sociales, los datos por lo general son de naturaleza no experimental, es decir, no están sujetos al control del investigador.[14] Por ejemplo, el investigador no ejerce ningún control directo sobre los datos del PIB, desempleo, precios de acciones, etc. Como veremos, esta falta de control a menudo genera problemas especiales para el investigador al identificar la causa o causas precisas que afectan una situación particular. Por ejemplo, ¿es la oferta monetaria la que determina el PIB (nominal) o la relación es inversa?

[12] Y. Grunfeld, "The Determinants of Corporate Investment", tesis doctoral inédita, Departamento de Economía, Universidad de Chicago, 1958. Estos datos son ya una herramienta excelente para ilustrar los modelos de regresión con datos en panel.

[13] Para mayor claridad, ver Albert T. Somers, *The U.S. Economy Demystified: What the Major Economic Statistics Mean and Their Significance for Business*, D.C. Heath, Lexington, Massachusetts, 1985.

[14] También en las ciencias sociales, con mucha frecuencia, hay experimentos controlados; se da un ejemplo en el ejercicio 1.6.

**TABLA 1.2 Datos de inversión de cuatro empresas, 1935-1954**

| Observación | $I$ | $F_{-1}$ | $C_{-1}$ | Observación | $I$ | $F_{-1}$ | $C_{-1}$ |
|---|---|---|---|---|---|---|---|
| **GE** | | | | **US** | | | |
| 1935 | 33.1 | 1 170.6 | 97.8 | 1935 | 209.9 | 1 362.4 | 53.8 |
| 1936 | 45.0 | 2 015.8 | 104.4 | 1936 | 355.3 | 1 807.1 | 50.5 |
| 1937 | 77.2 | 2 803.3 | 118.0 | 1937 | 469.9 | 2 673.3 | 118.1 |
| 1938 | 44.6 | 2 039.7 | 156.2 | 1938 | 262.3 | 1 801.9 | 260.2 |
| 1939 | 48.1 | 2 256.2 | 172.6 | 1939 | 230.4 | 1 957.3 | 312.7 |
| 1940 | 74.4 | 2 132.2 | 186.6 | 1940 | 361.6 | 2 202.9 | 254.2 |
| 1941 | 113.0 | 1 834.1 | 220.9 | 1941 | 472.8 | 2 380.5 | 261.4 |
| 1942 | 91.9 | 1 588.0 | 287.8 | 1942 | 445.6 | 2 168.6 | 298.7 |
| 1943 | 61.3 | 1 749.4 | 319.9 | 1943 | 361.6 | 1 985.1 | 301.8 |
| 1944 | 56.8 | 1 687.2 | 321.3 | 1944 | 288.2 | 1 813.9 | 279.1 |
| 1945 | 93.6 | 2 007.7 | 319.6 | 1945 | 258.7 | 1 850.2 | 213.8 |
| 1946 | 159.9 | 2 208.3 | 346.0 | 1946 | 420.3 | 2 067.7 | 232.6 |
| 1947 | 147.2 | 1 656.7 | 456.4 | 1947 | 420.5 | 1 796.7 | 264.8 |
| 1948 | 146.3 | 1 604.4 | 543.4 | 1948 | 494.5 | 1 625.8 | 306.9 |
| 1949 | 98.3 | 1 431.8 | 618.3 | 1949 | 405.1 | 1 667.0 | 351.1 |
| 1950 | 93.5 | 1 610.5 | 647.4 | 1950 | 418.8 | 1 677.4 | 357.8 |
| 1951 | 135.2 | 1 819.4 | 671.3 | 1951 | 588.2 | 2 289.5 | 341.1 |
| 1952 | 157.3 | 2 079.7 | 726.1 | 1952 | 645.2 | 2 159.4 | 444.2 |
| 1953 | 179.5 | 2 371.6 | 800.3 | 1953 | 641.0 | 2 031.3 | 623.6 |
| 1954 | 189.6 | 2 759.9 | 888.9 | 1954 | 459.3 | 2 115.5 | 669.7 |
| **GM** | | | | **WEST** | | | |
| 1935 | 317.6 | 3 078.5 | 2.8 | 1935 | 12.93 | 191.5 | 1.8 |
| 1936 | 391.8 | 4 661.7 | 52.6 | 1936 | 25.90 | 516.0 | 0.8 |
| 1937 | 410.6 | 5 387.1 | 156.9 | 1937 | 35.05 | 729.0 | 7.4 |
| 1938 | 257.7 | 2 792.2 | 209.2 | 1938 | 22.89 | 560.4 | 18.1 |
| 1939 | 330.8 | 4 313.2 | 203.4 | 1939 | 18.84 | 519.9 | 23.5 |
| 1940 | 461.2 | 4 643.9 | 207.2 | 1940 | 28.57 | 628.5 | 26.5 |
| 1941 | 512.0 | 4 551.2 | 255.2 | 1941 | 48.51 | 537.1 | 36.2 |
| 1942 | 448.0 | 3 244.1 | 303.7 | 1942 | 43.34 | 561.2 | 60.8 |
| 1943 | 499.6 | 4 053.7 | 264.1 | 1943 | 37.02 | 617.2 | 84.4 |
| 1944 | 547.5 | 4 379.3 | 201.6 | 1944 | 37.81 | 626.7 | 91.2 |
| 1945 | 561.2 | 4 840.9 | 265.0 | 1945 | 39.27 | 737.2 | 92.4 |
| 1946 | 688.1 | 4 900.0 | 402.2 | 1946 | 53.46 | 760.5 | 86.0 |
| 1947 | 568.9 | 3 526.5 | 761.5 | 1947 | 55.56 | 581.4 | 111.1 |
| 1948 | 529.2 | 3 245.7 | 922.4 | 1948 | 49.56 | 662.3 | 130.6 |
| 1949 | 555.1 | 3 700.2 | 1 020.1 | 1949 | 32.04 | 583.8 | 141.8 |
| 1950 | 642.9 | 3 755.6 | 1 099.0 | 1950 | 32.24 | 635.2 | 136.7 |
| 1951 | 755.9 | 4 833.0 | 1 207.7 | 1951 | 54.38 | 732.8 | 129.7 |
| 1952 | 891.2 | 4 924.9 | 1 430.5 | 1952 | 71.78 | 864.1 | 145.5 |
| 1953 | 1 304.4 | 6 241.7 | 1 777.3 | 1953 | 90.08 | 1 193.5 | 174.8 |
| 1954 | 1 486.7 | 5 593.6 | 2 226.3 | 1954 | 68.60 | 1 188.9 | 213.5 |

*Notas:* $Y = I$ = inversión bruta = adiciones a planta y equipo más mantenimiento y reparaciones, en millones de dólares deflacionados por $P_1$.

$X_2 = F$ = valor de la empresa = precio de las acciones comunes y preferentes al 31 de diciembre (o precio promedio del 31 de diciembre y 31 de enero del año siguiente) multiplicado por el número de acciones comunes y preferentes en circulación más el valor total de la deuda en libros al 31 de diciembre, en millones de dólares deflacionados por $P_2$.

$X_3 = C$ = existencias de planta y equipo = suma acumulada de adiciones netas a planta y equipo deflacionadas por $P_1$ menos provisión para depreciación deflacionada por $P_3$ en estas definiciones.

$P_1$ = factor de deflación implícito de los precios de equipo duradero de los productores (1947 = 100).

$P_2$ = factor de deflación implícito de los precios del PIB (1947 = 100).

$P_3$ = factor de deflación del gasto de depreciación = promedio móvil de 10 años del índice de precios al mayoreo de metales y productos metálicos (1947 = 100).

Fuente: Reproducido de H. D. Vinod y Aman Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, Nueva York, 1981, pp. 259-261.

## Precisión de los datos[15]

Si bien se dispone de numerosos datos para la investigación económica, su calidad no siempre es adecuada, y por múltiples razones.

1. Como ya vimos, en su mayoría, los datos de las ciencias sociales son de naturaleza no experimental. Por consiguiente, es posible incurrir en errores de observación, sea por acción u omisión.
2. Aun en datos reunidos experimentalmente surgen errores de medición debido a las aproximaciones o al redondeo.
3. En encuestas por cuestionarios, el problema de la falta de respuesta puede ser grave; un investigador tiene suerte si obtiene una tasa de respuesta de 40%. El análisis basado en dicha tasa de respuesta parcial quizá no refleje de verdad el comportamiento del 60% que no respondió, y ocasione, por consiguiente, un sesgo de selectividad (muestral). Además, existe el problema de quienes responden el cuestionario pero no todas las preguntas, sobre todo las que son delicadas por tratar cuestiones financieras, lo que genera un sesgo adicional de selectividad.
4. Los métodos de muestreo para obtención de datos llegan a variar tanto que a menudo es difícil comparar los resultados de las diversas muestras.
5. Las cifras económicas suelen estar disponibles en niveles muy agregados. Por ejemplo, la mayor parte de los macrodatos (como el PIB, empleo, inflación, desempleo) están disponibles para la economía en su conjunto, o, en el mejor de los casos, para algunas regiones geográficas muy amplias. Los datos con estos niveles tan elevados de agregación tal vez no ilustren mucho sobre los sujetos o las microunidades objeto de estudio.
6. Debido a su carácter confidencial, ciertos datos sólo pueden publicarse en forma muy agregada. En el caso de Estados Unidos, por ejemplo, la ley prohíbe al IRS (hacienda) revelar información sobre declaraciones de impuestos individuales; sólo puede revelar algunos datos generales. Por consiguiente, si se desea conocer el monto gastado en salud por los individuos con cierto nivel de ingresos, sólo es posible en un nivel muy agregado. Pero los macroanálisis de este tipo con frecuencia resultan insuficientes para revelar la dinámica del comportamiento de las microunidades. De igual forma, el Departamento de Comercio estadounidense, que levanta el censo de empresas cada cinco años, no tiene autorización para revelar información sobre producción, empleo, consumo de energía, gastos de investigación y desarrollo, etc., de las empresas. Así, es difícil estudiar las diferencias entre las empresas en estos aspectos.

Por estos problemas, y muchos más, **el investigador debe tener siempre en mente que el resultado de la investigación será tan bueno como lo sea la calidad de los datos**. Por tanto, si en algunas situaciones los investigadores concluyen que los resultados de la investigación son “insatisfactorios”, la causa puede ser la mala calidad de los datos y no un modelo equivocado. Por desgracia, debido a la naturaleza no experimental de los datos de la mayoría de los estudios de ciencias sociales, los investigadores con frecuencia no tienen más remedio que depender de la información disponible. Sin embargo, siempre deben tener presente que los datos pueden no ser los mejores y tratar de no ser muy dogmáticos sobre los resultados de un estudio dado, sobre todo cuando la calidad de los datos no es confiable.

## Una observación sobre las escalas de medición de las variables[16]

Las variables que a menudo encontrará se clasifican en cuatro categorías generales: *escala de razón*, *escala de intervalo*, *escala ordinal* y *escala nominal*. Es importante comprender cada una.

[15] Para un examen crítico, ver O. Morgenstern, *The Accuracy of Economic Observations*, 2a. ed., Princeton University Press, Princeton, Nueva Jersey, 1963.

[16] El siguiente análisis se basa en gran medida en Aris Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, Nueva York, 1999, p. 24.

### *Escala de razón*

Para la variable $X$, al tomar dos valores ($X_1$ y $X_2$), la razón $X_1/X_2$ y la distancia ($X_2 - X_1$) son cantidades con un significado. Asimismo, hay un ordenamiento natural (ascendente o descendente) de los valores a lo largo de la escala. En virtud de lo anterior, son sensatas las comparaciones como $X_2 \leq X_1$ o $X_2 \geq X_1$. En su mayoría, las variables económicas pertenecen a esta categoría. Por consiguiente, no es descabellado preguntar a cuánto asciende el PIB de este año en comparación con el del año anterior. El ingreso personal, en dólares, es una variable de razón; alguien que gana 100 000 dólares recibe el doble que quien percibe 50 000 (antes de impuestos, desde luego).

### *Escala de intervalo*

Una variable en escala de intervalo satisface las dos últimas propiedades de la variable en escala de razón, pero no la primera. Por tanto, la distancia entre dos periodos, (digamos 2000-1995), tiene significado, no así la razón de dos periodos (2000/1995). A las 11 de la mañana (hora de la costa del Pacífico de Estados Unidos) del 11 de agosto de 2007 se registró en Portland, Oregon, una temperatura de 60° Fahrenheit (15.5° Celsius), y en Tallahassee, Florida, de 90° F (32° C). La temperatura con esta escala no se mide en escala de razón pues no tiene sentido decir que en Tallahassee hizo 50% más calor que en Portland. Esto se debe sobre todo a que la escala Fahrenheit no usa 0° como base natural.

### *Escala ordinal*

Una variable pertenece a esta categoría sólo si satisface la tercera propiedad de la escala de razón (es decir, el orden natural), como los sistemas de calificaciones por letras (A, B, C) o los niveles de ingresos alto, medio y bajo). Para estas variables hay un orden, pero las distancias entre las categorías no son cuantificables. Los estudiantes de economía recordarán las *curvas de indiferencia* entre dos bienes, en donde una curva superior de indiferencia señala un mayor nivel de utilidad, pero no se puede cuantificar en qué medida una curva de indiferencia es mayor que otra.

### *Escala nominal*

Las variables de esta categoría no tienen ninguna característica de las variables en escala de razón. Las variables como el género (masculino y femenino) y el estado civil (casado, soltero, divorciado, separado) simplemente denotan categorías. *Pregunta:* ¿por qué no expresar dichas variables con las escalas de razón, intervalo u orden?

Como veremos, las técnicas econométricas adecuadas para las variables en escala de razón no resultarían pertinentes para las variables en escala nominal. En consecuencia, es importante tener en mente las diferencias entre los cuatro tipos de escalas de medición recién analizadas.

## Resumen y conclusiones

1. La idea fundamental del análisis de regresión es la dependencia estadística de una variable, la dependiente, respecto de otra o más variables, las explicativas.
2. El objetivo de tal análisis es estimar o predecir la media o el valor promedio de la variable dependiente con base en los valores conocidos o fijos de las explicativas.
3. En la práctica, un buen análisis de regresión depende de la disponibilidad de datos apropiados. En este capítulo analizamos la naturaleza, fuentes y limitaciones de los datos disponibles para la investigación, en especial en las ciencias sociales.
4. En toda investigación se debe señalar con claridad las fuentes de los datos para el análisis, sus definiciones, sus métodos de recolección y cualquier laguna u omisión en ellos, así como toda revisión que se les haya aplicado. Tenga en cuenta que los datos macroeconómicos que publica el gobierno con frecuencia son objeto de revisión.
5. Como el lector tal vez no tenga tiempo, energía o recursos para llegar a la fuente original de los datos, tiene el derecho de suponer que el investigador los recopiló de manera apropiada, y que los cálculos y análisis son correctos.

## EJERCICIOS

1.1. La tabla 1.3 proporciona datos sobre el índice de precios al consumidor de siete países industrializados, cuya base es 1982-1984 = 100.

*a*) A partir de estos datos, calcule la tasa de inflación en cada país.[17]

*b*) Grafique la tasa de inflación de cada nación en función del tiempo (es decir, asigne el eje horizontal al tiempo, y el vertical, a la tasa de inflación).

*c*) ¿Qué conclusiones generales surgen respecto de la inflación en los siete países?

*d*) ¿Qué país tiene, al parecer, la tasa de inflación más variable? ¿Puede explicarlo?

1.2. *a*) Use la tabla 1.3 para trazar la gráfica de la tasa de inflación de Canadá, Francia, Alemania, Italia, Japón y Gran Bretaña, en comparación con la de Estados Unidos.

*b*) Comente en términos generales el comportamiento de la tasa de inflación de los seis países, en comparación con la de Estados Unidos.

*c*) Si descubre que las tasas de inflación de esos seis países tienen la misma dirección que la de Estados Unidos, ¿esto indicaría que la inflación en Estados Unidos "provoca" la inflación en los demás países? ¿Por qué?

**TABLA 1.3**
**IPC de siete países industrializados, 1980-2005 (1982-1984 = 100)**

Fuente: *Economic Report of the President,* 2007, tabla 108, p. 354.

| Año | Estados Unidos | Canadá | Japón | Francia | Alemania | Italia | Gran Bretaña |
|---|---|---|---|---|---|---|---|
| 1980 | 82.4 | 76.1 | 91.0 | 72.2 | 86.7 | 63.9 | 78.5 |
| 1981 | 90.9 | 85.6 | 95.3 | 81.8 | 92.2 | 75.5 | 87.9 |
| 1982 | 96.5 | 94.9 | 98.1 | 91.7 | 97.0 | 87.8 | 95.4 |
| 1983 | 99.6 | 100.4 | 99.8 | 100.3 | 100.3 | 100.8 | 99.8 |
| 1984 | 103.9 | 104.7 | 102.1 | 108.0 | 102.7 | 111.4 | 104.8 |
| 1985 | 107.6 | 109.0 | 104.2 | 114.3 | 104.8 | 121.7 | 111.1 |
| 1986 | 109.6 | 113.5 | 104.9 | 117.2 | 104.6 | 128.9 | 114.9 |
| 1987 | 113.6 | 118.4 | 104.9 | 121.1 | 104.9 | 135.1 | 119.7 |
| 1988 | 118.3 | 123.2 | 105.6 | 124.3 | 106.3 | 141.9 | 125.6 |
| 1989 | 124.0 | 129.3 | 108.0 | 128.7 | 109.2 | 150.7 | 135.4 |
| 1990 | 130.7 | 135.5 | 111.4 | 132.9 | 112.2 | 160.4 | 148.2 |
| 1991 | 136.2 | 143.1 | 115.0 | 137.2 | 116.3 | 170.5 | 156.9 |
| 1992 | 140.3 | 145.3 | 117.0 | 140.4 | 122.2 | 179.5 | 162.7 |
| 1993 | 144.5 | 147.9 | 118.5 | 143.4 | 127.6 | 187.7 | 165.3 |
| 1994 | 148.2 | 148.2 | 119.3 | 145.8 | 131.1 | 195.3 | 169.3 |
| 1995 | 152.4 | 151.4 | 119.2 | 148.4 | 133.3 | 205.6 | 175.2 |
| 1996 | 156.9 | 153.8 | 119.3 | 151.4 | 135.3 | 213.8 | 179.4 |
| 1997 | 160.5 | 156.3 | 121.5 | 153.2 | 137.8 | 218.2 | 185.1 |
| 1998 | 163.0 | 157.8 | 122.2 | 154.2 | 139.1 | 222.5 | 191.4 |
| 1999 | 166.6 | 160.5 | 121.8 | 155.0 | 140.0 | 226.2 | 194.3 |
| 2000 | 172.2 | 164.9 | 121.0 | 157.6 | 142.0 | 231.9 | 200.1 |
| 2001 | 177.1 | 169.1 | 120.1 | 160.2 | 144.8 | 238.3 | 203.6 |
| 2002 | 179.9 | 172.9 | 119.0 | 163.3 | 146.7 | 244.3 | 207.0 |
| 2003 | 184.0 | 177.7 | 118.7 | 166.7 | 148.3 | 250.8 | 213.0 |
| 2004 | 188.9 | 181.0 | 118.7 | 170.3 | 150.8 | 256.3 | 219.4 |
| 2005 | 195.3 | 184.9 | 118.3 | 173.2 | 153.7 | 261.3 | 225.6 |

[17] Reste del IPC del año en cuestión el IPC del año anterior, divida la diferencia entre el IPC del año anterior y multiplique el resultado por 100. Así, la tasa de inflación de Canadá en 1981 fue de $[(85.6 - 76.1)/76.1] \times 100 = 12.48\%$ (aproximadamente).

1.3. En la tabla 1.4 se presentan los tipos de cambio de nueve países industrializados correspondientes a 1985-2006. Excepto para Gran Bretaña, el tipo de cambio se define como las unidades de la divisa equivalentes a un dólar estadounidense; para ese país, se define como el número de dólares estadounidenses que se cambian por una libra inglesa.

*a*) Grafique los tipos de cambio en función del tiempo y comente el comportamiento general de los tipos de cambio durante el periodo.

*b*) Se dice que el dólar tiene una *apreciación* si se compran más unidades de una divisa. Por el contrario, se dice que sufre una *depreciación* si se adquieren menos divisas. En el periodo 1985-2006, en general, ¿qué comportamiento tuvo el dólar estadounidense? A propósito, busque en un texto de macroeconomía o economía internacional los factores que determinan la apreciación o depreciación de una moneda.

1.4. En la tabla 1.5 aparecen los datos en los que se basa la oferta de dinero M1 de la figura 1.5. ¿Puede explicar por qué se incrementó la oferta de dinero durante el periodo que se muestra en la tabla?

1.5. Suponga que va a crear un modelo económico de actividades delictivas en el que considere las horas invertidas en ellas (por ejemplo, en la venta de drogas). ¿Qué variables tomaría en cuenta para crear dicho modelo? Vea si su modelo se asemeja al del economista ganador del premio Nobel, Gary Becker.[18]

**TABLA 1.4** **Tipos de cambio de nueve países: 1985-2006**

| Año | Australia | Canadá | R. P. de China | Japón | México | Corea del Sur | Suecia | Suiza | Gran Bretaña |
|---|---|---|---|---|---|---|---|---|---|
| 1985 | 0.7003 | 1.3659 | 2.9434 | 238.47 | 0.257 | 872.45 | 8.6032 | 2.4552 | 1.2974 |
| 1986 | 0.6709 | 1.3896 | 3.4616 | 168.35 | 0.612 | 884.60 | 7.1273 | 1.7979 | 1.4677 |
| 1987 | 0.7014 | 1.3259 | 3.7314 | 144.60 | 1.378 | 826.16 | 6.3469 | 1.4918 | 1.6398 |
| 1988 | 0.7841 | 1.2306 | 3.7314 | 128.17 | 2.273 | 734.52 | 6.1370 | 1.4643 | 1.7813 |
| 1989 | 0.7919 | 1.1842 | 3.7673 | 138.07 | 2.461 | 674.13 | 6.4559 | 1.6369 | 1.6382 |
| 1990 | 0.7807 | 1.1668 | 4.7921 | 145.00 | 2.813 | 710.64 | 5.9231 | 1.3901 | 1.7841 |
| 1991 | 0.7787 | 1.1460 | 5.3337 | 134.59 | 3.018 | 736.73 | 6.0521 | 1.4356 | 1.7674 |
| 1992 | 0.7352 | 1.2085 | 5.5206 | 126.78 | 3.095 | 784.66 | 5.8258 | 1.4064 | 1.7663 |
| 1993 | 0.6799 | 1.2902 | 5.7795 | 111.08 | 3.116 | 805.75 | 7.7956 | 1.4781 | 1.5016 |
| 1994 | 0.7316 | 1.3664 | 8.6397 | 102.18 | 3.385 | 806.93 | 7.7161 | 1.3667 | 1.5319 |
| 1995 | 0.7407 | 1.3725 | 8.3700 | 93.96 | 6.447 | 772.69 | 7.1406 | 1.1812 | 1.5785 |
| 1996 | 0.7828 | 1.3638 | 8.3389 | 108.78 | 7.600 | 805.00 | 6.7082 | 1.2361 | 1.5607 |
| 1997 | 0.7437 | 1.3849 | 8.3193 | 121.06 | 7.918 | 953.19 | 7.6446 | 1.4514 | 1.6376 |
| 1998 | 0.6291 | 1.4836 | 8.3008 | 130.99 | 9.152 | 1 400.40 | 7.9522 | 1.4506 | 1.6573 |
| 1999 | 0.6454 | 1.4858 | 8.2783 | 113.73 | 9.553 | 1 189.84 | 8.2740 | 1.5045 | 1.6172 |
| 2000 | 0.5815 | 1.4855 | 8.2784 | 107.80 | 9.459 | 1 130.90 | 9.1735 | 1.6904 | 1.5156 |
| 2001 | 0.5169 | 1.5487 | 8.2770 | 121.57 | 9.337 | 1 292.02 | 10.3425 | 1.6891 | 1.4396 |
| 2002 | 0.5437 | 1.5704 | 8.2771 | 125.22 | 9.663 | 1 250.31 | 9.7233 | 1.5567 | 1.5025 |
| 2003 | 0.6524 | 1.4008 | 8.2772 | 115.94 | 10.793 | 1 192.08 | 8.0787 | 1.3450 | 1.6347 |
| 2004 | 0.7365 | 1.3017 | 8.2768 | 108.15 | 11.290 | 1 145.24 | 7.3480 | 1.2428 | 1.8330 |
| 2005 | 0.7627 | 1.2115 | 8.1936 | 110.11 | 10.894 | 1 023.75 | 7.4710 | 1.2459 | 1.8204 |
| 2006 | 0.7535 | 1.1340 | 7.9723 | 116.31 | 10.906 | 954.32 | 7.3718 | 1.2532 | 1.8434 |

Fuente: *Economic Report of the President*, 2007, tabla B-110, p. 356.

[18] G. S. Becker, "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, vol. 76, 1968, pp. 169-217.

**TABLA 1.5**
**Oferta de dinero M1 ajustada por estacionalidad: enero de 1959-julio de 1999 (miles de millones de dólares)**

Fuente: Board of Governors, Federal Reserve Bank, Estados Unidos.

| | | | | | | |
|---|---|---|---|---|---|---|
| 1959:01 | 138.8900 | 139.3900 | 139.7400 | 139.6900 | 140.6800 | 141.1700 |
| 1959:07 | 141.7000 | 141.9000 | 141.0100 | 140.4700 | 140.3800 | 139.9500 |
| 1960:01 | 139.9800 | 139.8700 | 139.7500 | 139.5600 | 139.6100 | 139.5800 |
| 1960:07 | 140.1800 | 141.3100 | 141.1800 | 140.9200 | 140.8600 | 140.6900 |
| 1961:01 | 141.0600 | 141.6000 | 141.8700 | 142.1300 | 142.6600 | 142.8800 |
| 1961:07 | 142.9200 | 143.4900 | 143.7800 | 144.1400 | 144.7600 | 145.2000 |
| 1962:01 | 145.2400 | 145.6600 | 145.9600 | 146.4000 | 146.8400 | 146.5800 |
| 1962:07 | 146.4600 | 146.5700 | 146.3000 | 146.7100 | 147.2900 | 147.8200 |
| 1963:01 | 148.2600 | 148.9000 | 149.1700 | 149.7000 | 150.3900 | 150.4300 |
| 1963:07 | 151.3400 | 151.7800 | 151.9800 | 152.5500 | 153.6500 | 153.2900 |
| 1964:01 | 153.7400 | 154.3100 | 154.4800 | 154.7700 | 155.3300 | 155.6200 |
| 1964:07 | 156.8000 | 157.8200 | 158.7500 | 159.2400 | 159.9600 | 160.3000 |
| 1965:01 | 160.7100 | 160.9400 | 161.4700 | 162.0300 | 161.7000 | 162.1900 |
| 1965:07 | 163.0500 | 163.6800 | 164.8500 | 165.9700 | 166.7100 | 167.8500 |
| 1966:01 | 169.0800 | 169.6200 | 170.5100 | 171.8100 | 171.3300 | 171.5700 |
| 1966:07 | 170.3100 | 170.8100 | 171.9700 | 171.1600 | 171.3800 | 172.0300 |
| 1967:01 | 171.8600 | 172.9900 | 174.8100 | 174.1700 | 175.6800 | 177.0200 |
| 1967:07 | 178.1300 | 179.7100 | 180.6800 | 181.6400 | 182.3800 | 183.2600 |
| 1968:01 | 184.3300 | 184.7100 | 185.4700 | 186.6000 | 187.9900 | 189.4200 |
| 1968:07 | 190.4900 | 191.8400 | 192.7400 | 194.0200 | 196.0200 | 197.4100 |
| 1969:01 | 198.6900 | 199.3500 | 200.0200 | 200.7100 | 200.8100 | 201.2700 |
| 1969:07 | 201.6600 | 201.7300 | 202.1000 | 202.9000 | 203.5700 | 203.8800 |
| 1970:01 | 206.2200 | 205.0000 | 205.7500 | 206.7200 | 207.2200 | 207.5400 |
| 1970:07 | 207.9800 | 209.9300 | 211.8000 | 212.8800 | 213.6600 | 214.4100 |
| 1971:01 | 215.5400 | 217.4200 | 218.7700 | 220.0000 | 222.0200 | 223.4500 |
| 1971:07 | 224.8500 | 225.5800 | 226.4700 | 227.1600 | 227.7600 | 228.3200 |
| 1972:01 | 230.0900 | 232.3200 | 234.3000 | 235.5800 | 235.8900 | 236.6200 |
| 1972:07 | 238.7900 | 240.9300 | 243.1800 | 245.0200 | 246.4100 | 249.2500 |
| 1973:01 | 251.4700 | 252.1500 | 251.6700 | 252.7400 | 254.8900 | 256.6900 |
| 1973:07 | 257.5400 | 257.7600 | 257.8600 | 259.0400 | 260.9800 | 262.8800 |
| 1974:01 | 263.7600 | 265.3100 | 266.6800 | 267.2000 | 267.5600 | 268.4400 |
| 1974:07 | 269.2700 | 270.1200 | 271.0500 | 272.3500 | 273.7100 | 274.2000 |
| 1975:01 | 273.9000 | 275.0000 | 276.4200 | 276.1700 | 279.2000 | 282.4300 |
| 1975:07 | 283.6800 | 284.1500 | 285.6900 | 285.3900 | 286.8300 | 287.0700 |
| 1976:01 | 288.4200 | 290.7600 | 292.7000 | 294.6600 | 295.9300 | 296.1600 |
| 1976:07 | 297.2000 | 299.0500 | 299.6700 | 302.0400 | 303.5900 | 306.2500 |
| 1977:01 | 308.2600 | 311.5400 | 313.9400 | 316.0200 | 317.1900 | 318.7100 |
| 1977:07 | 320.1900 | 322.2700 | 324.4800 | 326.4000 | 328.6400 | 330.8700 |
| 1978:01 | 334.4000 | 335.3000 | 336.9600 | 339.9200 | 344.8600 | 346.8000 |
| 1978:07 | 347.6300 | 349.6600 | 352.2600 | 353.3500 | 355.4100 | 357.2800 |
| 1979:01 | 358.6000 | 359.9100 | 362.4500 | 368.0500 | 369.5900 | 373.3400 |
| 1979:07 | 377.2100 | 378.8200 | 379.2800 | 380.8700 | 380.8100 | 381.7700 |
| 1980:01 | 385.8500 | 389.7000 | 388.1300 | 383.4400 | 384.6000 | 389.4600 |
| 1980:07 | 394.9100 | 400.0600 | 405.3600 | 409.0600 | 410.3700 | 408.0600 |
| 1981:01 | 410.8300 | 414.3800 | 418.6900 | 427.0600 | 424.4300 | 425.5000 |
| 1981:07 | 427.9000 | 427.8500 | 427.4600 | 428.4500 | 430.8800 | 436.1700 |
| 1982:01 | 442.1300 | 441.4900 | 442.3700 | 446.7800 | 446.5300 | 447.8900 |
| 1982:07 | 449.0900 | 452.4900 | 457.5000 | 464.5700 | 471.1200 | 474.3000 |
| 1983:01 | 476.6800 | 483.8500 | 490.1800 | 492.7700 | 499.7800 | 504.3500 |
| 1983:07 | 508.9600 | 511.6000 | 513.4100 | 517.2100 | 518.5300 | 520.7900 |
| 1984:01 | 524.4000 | 526.9900 | 530.7800 | 534.0300 | 536.5900 | 540.5400 |
| 1984:07 | 542.1300 | 542.3900 | 543.8600 | 543.8700 | 547.3200 | 551.1900 |
| 1985:01 | 555.6600 | 562.4800 | 565.7400 | 569.5500 | 575.0700 | 583.1700 |
| 1985:07 | 590.8200 | 598.0600 | 604.4700 | 607.9100 | 611.8300 | 619.3600 |
| 1986:01 | 620.4000 | 624.1400 | 632.8100 | 640.3500 | 652.0100 | 661.5200 |

(*continúa*)

**TABLA 1.5** (*continuación*)

| | | | | | | |
|---|---|---|---|---|---|---|
| 1986:07 | 672.2000 | 680.7700 | 688.5100 | 695.2600 | 705.2400 | 724.2800 |
| 1987:01 | 729.3400 | 729.8400 | 733.0100 | 743.3900 | 746.0000 | 743.7200 |
| 1987:07 | 744.9600 | 746.9600 | 748.6600 | 756.5000 | 752.8300 | 749.6800 |
| 1988:01 | 755.5500 | 757.0700 | 761.1800 | 767.5700 | 771.6800 | 779.1000 |
| 1988:07 | 783.4000 | 785.0800 | 784.8200 | 783.6300 | 784.4600 | 786.2600 |
| 1989:01 | 784.9200 | 783.4000 | 782.7400 | 778.8200 | 774.7900 | 774.2200 |
| 1989:07 | 779.7100 | 781.1400 | 782.2000 | 787.0500 | 787.9500 | 792.5700 |
| 1990:01 | 794.9300 | 797.6500 | 801.2500 | 806.2400 | 804.3600 | 810.3300 |
| 1990:07 | 811.8000 | 817.8500 | 821.8300 | 820.3000 | 822.0600 | 824.5600 |
| 1991:01 | 826.7300 | 832.4000 | 838.6200 | 842.7300 | 848.9600 | 858.3300 |
| 1991:07 | 862.9500 | 868.6500 | 871.5600 | 878.4000 | 887.9500 | 896.7000 |
| 1992:01 | 910.4900 | 925.1300 | 936.0000 | 943.8900 | 950.7800 | 954.7100 |
| 1992:07 | 964.6000 | 975.7100 | 988.8400 | 1 004.340 | 1 016.040 | 1 024.450 |
| 1993:01 | 1 030.900 | 1 033.150 | 1 037.990 | 1 047.470 | 1 066.220 | 1 075.610 |
| 1993:07 | 1 085.880 | 1 095.560 | 1 105.430 | 1 113.800 | 1 123.900 | 1 129.310 |
| 1994:01 | 1 132.200 | 1 136.130 | 1 139.910 | 1 141.420 | 1 142.850 | 1 145.650 |
| 1994:07 | 1 151.490 | 1 151.390 | 1 152.440 | 1 150.410 | 1 150.440 | 1 149.750 |
| 1995:01 | 1 150.640 | 1 146.740 | 1 146.520 | 1 149.480 | 1 144.650 | 1 144.240 |
| 1995:07 | 1 146.500 | 1 146.100 | 1 142.270 | 1 136.430 | 1 133.550 | 1 126.730 |
| 1996:01 | 1 122.580 | 1 117.530 | 1 122.590 | 1 124.520 | 1 116.300 | 1 115.470 |
| 1996:07 | 1 112.340 | 1 102.180 | 1 095.610 | 1 082.560 | 1 080.490 | 1 081.340 |
| 1997:01 | 1 080.520 | 1 076.200 | 1 072.420 | 1 067.450 | 1 063.370 | 1 065.990 |
| 1997:07 | 1 067.570 | 1 072.080 | 1 064.820 | 1 062.060 | 1 067.530 | 1 074.870 |
| 1998:01 | 1 073.810 | 1 076.020 | 1 080.650 | 1 082.090 | 1 078.170 | 1 077.780 |
| 1998:07 | 1 075.370 | 1 072.210 | 1 074.650 | 1 080.400 | 1 088.960 | 1 093.350 |
| 1999:01 | 1 091.000 | 1 092.650 | 1 102.010 | 1 108.400 | 1 104.750 | 1 101.110 |
| 1999:07 | 1 099.530 | 1 102.400 | 1 093.460 | | | |

1.6. *Experimentos controlados en economía:* El 7 de abril de 2000, el presidente Clinton convirtió en ley una propuesta aprobada por ambas cámaras legislativas estadounidenses mediante la cual se eliminaban las limitaciones de beneficios para los derechohabientes del sistema de seguridad social. Antes de esa ley, los derechohabientes de entre 65 y 69 años con percepciones mayores de 17 000 dólares al año perdían el equivalente a un dólar de las prestaciones de seguridad social por cada 3 dólares de ingresos que excedieran 17 000 dólares. ¿Cómo se planearía un estudio que evaluara el efecto de este cambio en la ley? *Nota:* En la ley derogada no había restricción de ingresos para los derechohabientes de más de 70 años.

1.7. Los datos de la tabla 1.6 se publicaron el primero de marzo de 1984 en el periódico *The Wall Street Journal*. Se refieren al presupuesto publicitario (en millones de dólares) de 21 empresas en 1983 y a los millones de impactos semanales (veces que los clientes ven los anuncios de los productos de dichas compañías por semana). La información se basa en una encuesta a 4 000 adultos en la que se pidió a los usuarios de los productos que mencionaran un comercial que hubieran visto en la semana anterior y que tuviera que ver con la categoría del producto.

*a*) Trace una gráfica con los impactos en el eje vertical y el gasto publicitario en el horizontal.

*b*) ¿Qué se puede decir sobre la relación entre ambas variables?

*c*) Al observar la gráfica, ¿cree que es redituable el gasto en publicidad? Piense en todos los comerciales que se transmiten el domingo que se juega el Super Bowl o durante la Serie Mundial del beisbol estadounidense.

*Nota:* En los siguientes capítulos estudiaremos más a fondo los datos de la tabla 1.6.

**TABLA 1.6**
**Efecto del gasto en publicidad**

Fuente: http://lib.stat.cmu.edu/DASL/Datafiles/tvadsdat.html.

| Empresa | Impactos, millones | Gasto, millones de dólares de 1983 |
|---|---|---|
| 1. Miller Lite | 32.1 | 50.1 |
| 2. Pepsi | 99.6 | 74.1 |
| 3. Stroh's | 11.7 | 19.3 |
| 4. Fed'l Express | 21.9 | 22.9 |
| 5. Burger King | 60.8 | 82.4 |
| 6. Coca-Cola | 78.6 | 40.1 |
| 7. McDonald's | 92.4 | 185.9 |
| 8. MCI | 50.7 | 26.9 |
| 9. Diet Cola | 21.4 | 20.4 |
| 10. Ford | 40.1 | 166.2 |
| 11. Levi's | 40.8 | 27.0 |
| 12. Bud Lite | 10.4 | 45.6 |
| 13. ATT/Bell | 88.9 | 154.9 |
| 14. Calvin Klein | 12.0 | 5.0 |
| 15. Wendy's | 29.2 | 49.7 |
| 16. Polaroid | 38.0 | 26.9 |
| 17. Shasta | 10.0 | 5.7 |
| 18. Meow Mix | 12.3 | 7.6 |
| 19. Oscar Meyer | 23.4 | 9.2 |
| 20. Crest | 71.1 | 32.4 |
| 21. Kibbles 'N Bits | 4.4 | 6.1 |

Capítulo 2

# Análisis de regresión con dos variables: algunas ideas básicas

En el capítulo 1 vimos el concepto de regresión en términos generales. En este capítulo lo abordamos más formalmente. De manera específica, este capítulo y los tres siguientes introducirán al lector a la teoría básica del análisis de regresión más sencillo posible, es decir, la regresión **bivariable** o con **dos variables**, en la cual la variable dependiente (la regresada) se relaciona con una sola variable explicativa (la regresora). Consideremos primero este caso no necesariamente por su adecuación práctica, sino porque presenta las ideas fundamentales del análisis de regresión de la manera más sencilla posible, y algunas de estas ideas pueden ilustrarse con diagramas bidimensionales. Además, como veremos, el análisis de regresión **múltiple**, más general, en el que la regresada se relaciona con más de una regresora, es, en muchos sentidos, una extensión lógica del caso de dos variables.

## 2.1 Ejemplo hipotético[1]

Como se señaló en la sección 1.2, el análisis de regresión se relaciona en gran medida con la estimación o predicción de la media (de la población) o valor promedio de la variable dependiente, con base en los valores conocidos o fijos de las variables explicativas.[2] Para entender esto, consideremos los datos de la tabla 2.1. Estos datos se refieren a la **población** total de 60 familias de una comunidad hipotética, así como a su ingreso semanal ($X$) y su gasto de consumo semanal ($Y$), en dólares. Las 60 familias se dividen en 10 grupos de ingresos (de 80 dólares a 260); asimismo, aparecen los gastos semanales de cada familia de los diversos grupos. Por consiguiente, hay 10 valores *fijos* de $X$ y los correspondientes valores $Y$ para cada valor $X$; así, hay 10 subpoblaciones $Y$.

Se observa una variación considerable en el consumo semanal de cada grupo de ingreso, lo cual se aprecia con claridad en la figura 2.1. No obstante, el panorama general es que, a pesar de la variabilidad del consumo semanal en cada nivel de ingreso considerado, *en promedio,* el consumo semanal se incrementa a medida que aumenta el ingreso. Para verificar lo anterior, en la

[1] Se recomienda al lector cuyos conocimientos de estadística estén un poco empolvados que, para refrescarlos, antes de leer este capítulo consulte el apéndice estadístico, **apéndice A**.

[2] El valor esperado, esperanza o media poblacional de una variable aleatoria $Y$ se denota con el símbolo $E(Y)$. Por otra parte, el valor medio calculado a partir de una muestra de valores tomada de una población $Y$ se denota como $\bar{Y}$, y se lee "Y barra".

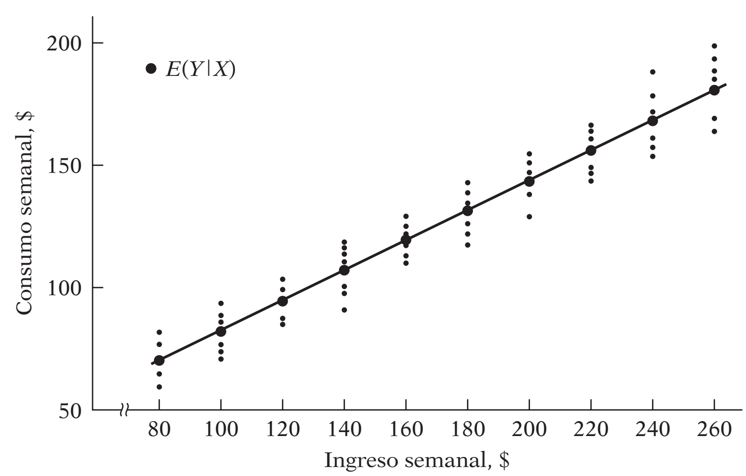
**TABLA 2.1**
**Ingreso familiar semanal *X*, $**

| Y↓ \ X→ | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
|---|---|---|---|---|---|---|---|---|---|---|
| Consumo familiar semanal $Y$, $ | 55 | 65 | 79 | 80 | 102 | 110 | 120 | 135 | 137 | 150 |
| | 60 | 70 | 84 | 93 | 107 | 115 | 136 | 137 | 145 | 152 |
| | 65 | 74 | 90 | 95 | 110 | 120 | 140 | 140 | 155 | 175 |
| | 70 | 80 | 94 | 103 | 116 | 130 | 144 | 152 | 165 | 178 |
| | 75 | 85 | 98 | 108 | 118 | 135 | 145 | 157 | 175 | 180 |
| | – | 88 | – | 113 | 125 | 140 | – | 160 | 189 | 185 |
| | – | – | – | 115 | – | – | – | 162 | – | 191 |
| Total | 325 | 462 | 445 | 707 | 678 | 750 | 685 | 1 043 | 966 | 1 211 |
| Media condicional de $Y$, $E(Y\|X)$ | 65 | 77 | 89 | 101 | 113 | 125 | 137 | 149 | 161 | 173 |

tabla 2.1 se proporciona la media, o promedio, del consumo semanal que corresponde a cada uno de los 10 niveles de ingresos. Así, al nivel de ingreso de 80 dólares le corresponde una media de consumo igual a 65 dólares, pero para el nivel de 200, la media es de 137. En total hay 10 valores medios para las 10 subpoblaciones de $Y$. A estos valores medios se les llama **valores esperados condicionales**, en virtud de que dependen de los valores de la variable (condicional) $X$. En forma simbólica, se denotan con $E(Y \mid X)$, lo cual se lee como el valor esperado de $Y$, dado el valor de $X$ (ver también la tabla 2.2).

Es importante distinguir entre los valores esperados condicionales y el **valor esperado incondicional** del consumo semanal, $E(Y)$. Si sumamos los consumos semanales de las 60 familias que forman la *población* y dividimos este número entre 60, obtendremos la cantidad de 121.20 dólares ($7 272/60), que es el valor de la media incondicional, o esperada, del consumo semanal, $E(Y)$; es incondicional porque, para obtener esta cifra, obviamos los niveles de ingreso de las diversas familias.[3] Como es lógico, los diferentes valores esperados condicionales de $Y$ de la

**FIGURA 2.1**
Distribución condicional del gasto en varios niveles de ingreso (datos de la tabla 2.1).




[3] Como se muestra en el **apéndice A**, los valores de las medias condicionales e incondicionales suelen diferir.

**TABLA 2.2**
**Probabilidades condicionales $p(Y \mid X_i)$ para los datos de la tabla 2.1**

| $p(Y \mid X_i)$ ↓ / $X$→ | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 260 |
|---|---|---|---|---|---|---|---|---|---|---|
| Probabilidades condicionales $p(Y \mid X_i)$ | $\frac{1}{5}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | $\frac{1}{5}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | $\frac{1}{5}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | $\frac{1}{5}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | $\frac{1}{5}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | — | $\frac{1}{6}$ | — | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | — | $\frac{1}{7}$ | $\frac{1}{6}$ | $\frac{1}{7}$ |
| | — | — | — | $\frac{1}{7}$ | — | — | — | $\frac{1}{7}$ | — | $\frac{1}{7}$ |
| Medias condicionales de $Y$ | 65 | 77 | 89 | 101 | 113 | 125 | 137 | 149 | 161 | 173 |

tabla 2.1 varían respecto del valor esperado incondicional de $Y$, igual a 121.20 dólares. Cuando se plantea la pregunta "¿cuál es el *valor esperado* del consumo semanal de una familia?", la respuesta es 121.20 dólares (la media incondicional). Pero si se pregunta "¿cuál es el *valor esperado* del consumo semanal de una familia cuyo ingreso mensual es de 140 dólares?", la respuesta es 101 (la media condicional). En otras palabras, a la pregunta "¿cuál es la mejor predicción (media) del gasto semanal de las familias con un ingreso semanal de 140 dólares?", la respuesta es 101 dólares. Por consiguiente, conocer el nivel de ingreso permite predecir mejor el valor medio del consumo que si se ignora esa información.[4] Tal vez sea ésta la esencia del análisis de regresión, como lo descubriremos a lo largo de este libro.

Los puntos oscuros dentro de círculos de la figura 2.1 muestran los valores medios condicionales de $Y$, graficados en función de los diferentes valores de $X$. Al unir esos valores obtenemos la **línea de regresión poblacional (LRP)**, o, más general, la **curva de regresión poblacional (CRP)**.[5] Con palabras más sencillas, es la **regresión de *Y* sobre *X***. El adjetivo "poblacional" se debe a que en este ejemplo trabajamos con la población total de 60 familias. Por supuesto, en realidad una población tendría más familias.

*Así, desde el punto de vista geométrico, una curva de regresión poblacional es tan sólo el lugar geométrico de las medias condicionales de la variable dependiente para los valores fijos de la(s) variable(s) explicativa(s).* En palabras más simples, es la curva que conecta las medias de las subpoblaciones de $Y$ que corresponden a los valores dados de la regresora $X$. Lo anterior se ve de manera gráfica en la figura 2.2.

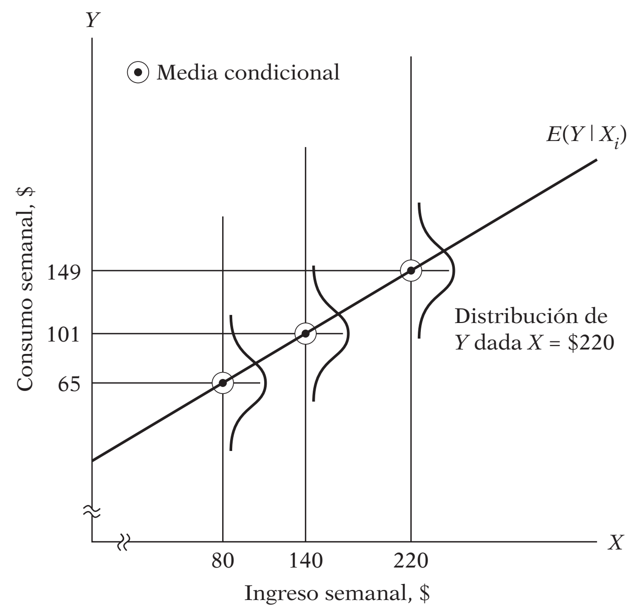
Esta figura muestra que para cada $X$ (es decir, el nivel de ingresos) existe una población de valores $Y$ (consumo semanal) que se distribuyen alrededor de la media (condicional) de dichos valores $Y$. Por simplicidad, suponemos que tales valores $Y$ están distribuidos simétricamente alrededor de sus respectivos valores medios (condicionales). Asimismo, la recta (o curva) de regresión pasa a través de los mencionados valores medios (condicionales).

Con estos antecedentes, es útil para el lector leer de nuevo la definición de regresión de la sección 1.2.

[4] Estoy en deuda con James Davidson por esta perspectiva. Ver James Davidson, *Econometric Theory*, Blackwell, Oxford, Gran Bretaña, 2000, p. 11.

[5] En el presente ejemplo, la LRP es una recta, pero podría ser una curva (ver la figura 2.3).

**FIGURA 2.2**
Línea de regresión poblacional (datos de la tabla 2.1).




## 2.2 Concepto de función de regresión poblacional (FRP)

De la anterior exposición, y en especial de las figuras 2.1 y 2.2, es claro que cada media condicional $E(Y \mid X_i)$ es función de $X_i$, donde $X_i$ es un valor dado de $X$. Simbólicamente,

$$E(Y \mid X_i) = f(X_i) \tag{2.2.1}$$

donde $f(X_i)$ denota alguna función de la variable explicativa $X$. En el ejemplo, $E(Y \mid X_i)$ es una función lineal de $X_i$. La ecuación (2.2.1) se conoce como **función de esperanza condicional (FEC), función de regresión poblacional (FRP)** o **regresión poblacional (RP)**, para abreviar. Dicha función sólo denota que el *valor esperado* de la distribución de $Y$ dada $X_i$ se relaciona funcionalmente con $X_i$. En otras palabras, dice cómo la media o respuesta promedio de $Y$ varía con $X$.

¿Qué forma adopta la función $f(X_i)$? Esta pregunta es importante porque en una situación real no disponemos de toda la población para efectuar el análisis. La forma funcional de la FRP es por consiguiente una pregunta empírica, aunque en casos específicos la teoría tiene algo que decir. Por ejemplo, un economista puede plantear que el consumo manifiesta una relación lineal con el ingreso. Por tanto, como primera aproximación o hipótesis de trabajo, podemos suponer que la FRP $E(Y \mid X_i)$ es una función lineal de $X_i$, del tipo

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i \tag{2.2.2}$$

donde $\beta_1$ y $\beta_2$ son parámetros no conocidos pero fijos que se denominan **coeficientes de regresión**; $\beta_1$ y $\beta_2$ se conocen también como **coeficientes de intersección** y **de pendiente**, respectivamente. La ecuación (2.2.1) se conoce como **función de regresión poblacional lineal**. En la bibliografía aparecen otras expresiones, como *modelo de regresión poblacional lineal* o sólo *regresión poblacional lineal*. En lo sucesivo, consideraremos sinónimos los términos **regresión**, **ecuación de regresión** y **modelo de regresión**.

En el análisis de regresión, la idea es estimar las FRP como la ecuación (2.2.2); es decir, estimar los valores no conocidos de $\beta_1$ y $\beta_2$ con base en las observaciones de $Y$ y $X$. Veremos este tema con más detalle en el capítulo 3.

## 2.3 Significado del término *lineal*

Como este texto se relaciona sobre todo con modelos lineales, como la ecuación (2.2.2), es esencial entender a cabalidad el término *lineal*, pues se interpreta de dos formas.

### Linealidad en las variables

El primer significado, y tal vez el más "natural", de linealidad es aquel en que la esperanza condicional de $Y$ es una función lineal de $X_i$, como en la ecuación (2.2.2).[6] Geométricamente, la curva de regresión en este caso es una recta. En esta interpretación, una función de regresión como $E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$ no es una función lineal porque la variable $X$ aparece elevada a una potencia o índice de 2.

### Linealidad en los parámetros

La segunda interpretación de linealidad se presenta cuando la esperanza condicional de $Y$, $E(Y \mid X_i)$, es una función lineal de los parámetros, los $\beta$; puede ser o no lineal en la variable $X$.[7] De acuerdo con esta interpretación, $E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$ es un modelo de regresión lineal (en el parámetro). Para ver lo anterior, supongamos que $X$ tiene un valor de 3. Por tanto, $E(Y \mid X = 3) = \beta_1 + 9\beta_2$, ecuación a todas luces lineal en $\beta_1$ y $\beta_2$. En consecuencia, todos los modelos de la figura 2.3 son de regresión lineal; es decir, son modelos lineales en los parámetros.

Ahora consideremos el modelo $E(Y \mid X_i) = \beta_1 + \beta_2^2 X_i$. Supongamos también que $X = 3$; así obtenemos $E(Y \mid X_i) = \beta_1 + 3\beta_2^2$, que es no lineal respecto del parámetro $\beta_2$. El anterior es un ejemplo de **modelo de regresión no lineal (en el parámetro)**. Analizaremos dichos modelos en el capítulo 14.

De las dos interpretaciones de linealidad, la linealidad en los parámetros es pertinente para el desarrollo de la teoría de regresión que presentaremos en breve. Por consiguiente, *en adelante, el término regresión "lineal" siempre significará una regresión lineal en los parámetros; los $\beta$ (es decir, los parámetros) se elevan sólo a la primera potencia. Puede o no ser lineal en las variables explicativas $X$.* Esquemáticamente tenemos la tabla 2.3. Así, $E(Y \mid X_i) = \beta_1 + \beta_2 X_i$, lineal en los parámetros igual que en las variables, es un modelo de regresión lineal (MRL), lo mismo que $E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$, lineal en los parámetros pero no lineal en la variable $X$.

[6] Se dice que una función $Y = f(X)$ es lineal en $X$ si $X$ aparece elevado a una potencia o índice de 1 solamente (es decir, se excluyen términos como $X^2$, $\sqrt{X}$ y demás), y dicha variable no está multiplicada ni dividida por alguna otra variable (por ejemplo, $X \cdot Z$ o $X/Z$, donde $Z$ es otra variable). Si $Y$ sólo depende de $X$, otra forma de plantear que $Y$ se relaciona linealmente con $X$ es que la tasa de cambio de $Y$ respecto de $X$ (es decir, la pendiente, o la derivada, de $Y$ respecto de $X$, $dY/dX$) es independiente del valor de $X$. Así, si $Y = 4X$, $dY/dX = 4$, valor independiente del valor de $X$. Pero si $Y = 4X^2$, $dY/dX = 8X$, término que no es independiente del valor tomado por $X$. Por consiguiente, la función no es lineal en $X$.

[7] Se dice que una función es lineal en el parámetro, digamos $\beta_1$, si $\beta_1$ aparece elevado a una potencia de 1 solamente y no está multiplicado ni dividido por ningún otro parámetro (por ejemplo, $\beta_1\beta_2$, $\beta_2/\beta_1$, etcétera).

**FIGURA 2.3**
Funciones lineales en los parámetros.

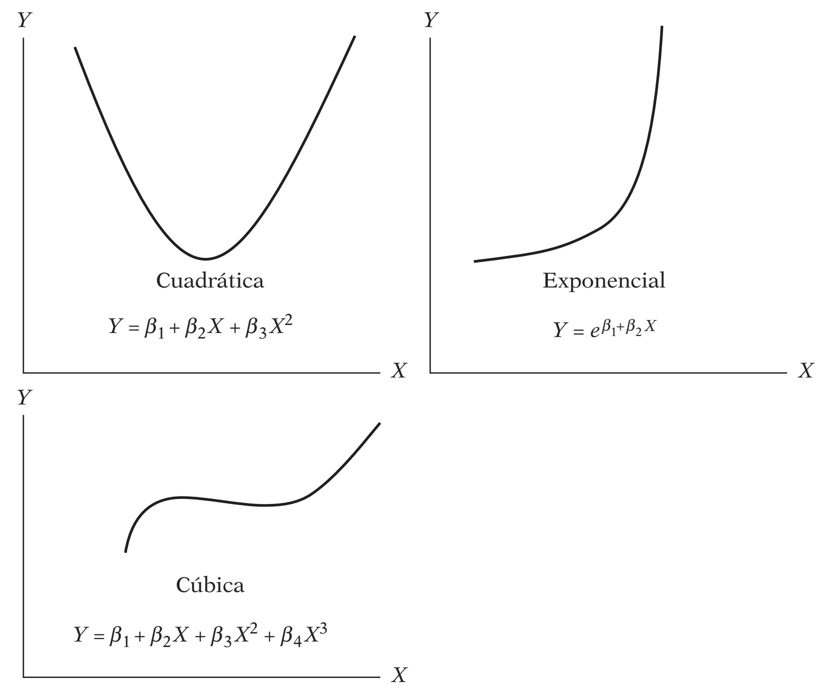



**TABLA 2.3**
**Modelos de regresión lineal**

| ¿Modelo lineal en los parámetros? | ¿Modelo lineal en las variables? | |
|---|---|---|
| | **Sí** | **No** |
| Sí | MRL | MRL |
| No | MRNL | MRNL |

*Nota:* MRL = modelo de regresión lineal
MRNL = modelo de regresión no lineal

## 2.4 Especificación estocástica de la FRP

En la figura 2.1 es claro que, a medida que aumenta el ingreso familiar, el consumo familiar, en promedio, también aumenta. Pero, ¿qué sucede con el consumo de una familia en relación con su nivel de ingreso (fijo)? Es obvio, por la tabla 2.1 y la figura 2.1, que el consumo de una familia en particular no necesariamente aumenta a medida que lo hace el nivel de ingreso. Por ejemplo, en la tabla 2.1 se observa que en el nivel de ingreso de 100 dólares existe una familia cuyo consumo, de 65, es menor que el consumo de dos familias cuyo ingreso semanal es sólo de 80 dólares. Sin embargo, hay que advertir que el consumo *promedio* de las familias con ingreso semanal de 100 es mayor que el consumo promedio de las familias con un ingreso semanal de 80 dólares (77 y 65).

Entonces, ¿qué se puede decir sobre la relación entre el consumo de una familia y un nivel determinado de ingresos? En la figura 2.1 se ve que, con el nivel de ingresos de $X_i$, el consumo

de una familia en particular se agrupa alrededor del consumo promedio de todas las familias en ese nivel de $X_i$, es decir, alrededor de su esperanza condicional. Por consiguiente, expresamos la *desviación* de un $Y_i$ en particular alrededor de su valor esperado de la manera siguiente:

$$u_i = Y_i - E(Y \mid X_i)$$

o

$$Y_i = E(Y \mid X_i) + u_i \tag{2.4.1}$$

donde la desviación $u_i$ es una variable aleatoria no observable que adopta valores positivos o negativos. Técnicamente, $u_i$ se conoce como **perturbación estocástica** o **término de error estocástico**.

¿Cómo se interpreta la ecuación (2.4.1)? Se puede decir que el gasto de una familia en particular, según su nivel de ingreso, se expresa como la suma de dos componentes: 1) $E(Y \mid X_i)$, que es simplemente la media del consumo de todas las familias con el mismo nivel de ingreso. Este componente se conoce como componente **sistemático**, o **determinista**, y 2) $u_i$ que es el componente aleatorio, o **no sistemático**. Examinaremos en breve la naturaleza del término de perturbación estocástica, pero por el momento supondremos que es un término *que sustituye* o *representa* a todas las variables omitidas o ignoradas que puedan afectar a $Y$ pero que no se incluyen (o no pueden incluirse) en el modelo de regresión.

Si suponemos que $E(Y \mid X_i)$ es lineal en $X_i$, como en (2.2.2), la ecuación (2.4.1) se escribe como

$$\begin{aligned} Y_i &= E(Y \mid X_i) + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned} \tag{2.4.2}$$

La ecuación (2.4.2) plantea que el consumo de una familia se relaciona linealmente con su ingreso más el término de perturbación. Así, el consumo individual, con $X = 80$ (ver la tabla 2.1), se expresa como

$$\begin{aligned} Y_1 &= 55 = \beta_1 + \beta_2(80) + u_1 \\ Y_2 &= 60 = \beta_1 + \beta_2(80) + u_2 \\ Y_3 &= 65 = \beta_1 + \beta_2(80) + u_3 \\ Y_4 &= 70 = \beta_1 + \beta_2(80) + u_4 \\ Y_5 &= 75 = \beta_1 + \beta_2(80) + u_5 \end{aligned} \tag{2.4.3}$$

Ahora, si tomamos el valor esperado de (2.4.1) en ambos lados, obtenemos

$$\begin{aligned} E(Y_i \mid X_i) &= E[E(Y \mid X_i)] + E(u_i \mid X_i) \\ &= E(Y \mid X_i) + E(u_i \mid X_i) \end{aligned} \tag{2.4.4}$$

donde se aprovecha que el valor esperado de una constante sea la constante misma.[8] Observe con atención que en la ecuación (2.4.4) se tomó la esperanza condicional, condicionada a las $X$ dadas.

Como $E(Y_i \mid X_i)$ es lo mismo que $E(Y \mid X_i)$, la ecuación (2.4.4) implica que

$$E(u_i \mid X_i) = 0 \tag{2.4.5}$$

[8] Ver en el **apéndice A** un breve análisis de las propiedades del operador de esperanza $E$. Observe que $E(Y \mid X_i)$, una vez fijo el valor de $X_i$, es una constante.

Así, el supuesto de que la línea de regresión pasa a través de las medias condicionales de $Y$ (ver la figura 2.2) implica que los valores de la media condicional de $u_i$ (condicionados al valor dado de $X$) son cero.

De la exposición anterior es claro que (2.2.2) y (2.4.2) son formas equivalentes si $E(u_i \mid X_i) = 0$.[9] Pero la especificación estocástica (2.4.2) tiene la ventaja que muestra claramente otras variables, además del ingreso, que afectan el consumo, y que el consumo de una familia no se explica en su totalidad sólo por la(s) variable(s) en el modelo de regresión.

## 2.5 Importancia del término de perturbación estocástica

Como anotamos en la sección 2.4, el término de perturbación $u_i$ es un sustituto de todas las variables que se omiten en el modelo, pero que, en conjunto, afectan a $Y$. La pregunta obvia es: ¿por qué no se introducen explícitamente estas variables en el modelo? O de otra forma, ¿por qué no se crea un modelo de regresión múltiple con tantas variables como sea posible? Las razones son muchas.

1. *Vaguedad de la teoría:* De existir una teoría que determine el comportamiento de $Y$, podría estar incompleta, y con frecuencia lo está. Se tendría quizá la certeza de que el ingreso semanal $X$ afecta el consumo semanal $Y$, pero también ignoraríamos, o no tendríamos la seguridad, sobre las demás variables que afectan a $Y$. Por consiguiente, $u_i$ sirve como sustituto de todas las variables excluidas u omitidas del modelo.

2. *Falta de disponibilidad de datos:* Aunque se conozcan algunas variables excluidas y se considerara por tanto una regresión múltiple en lugar de una simple, tal vez no se cuente con información cuantitativa sobre esas variables. Es común en el análisis empírico que no se disponga de los datos que idealmente se desearía tener. Por ejemplo, en principio se puede introducir la riqueza familiar como variable explicativa adicional a la variable ingreso para explicar el consumo familiar. Pero, por desgracia, la información sobre riqueza familiar por lo general no está disponible. Así, no habría mas que omitir la variable riqueza del modelo a pesar de su gran relevancia teórica para explicar el consumo.

3. *Variables centrales y variables periféricas:* Suponga en el ejemplo consumo-ingreso que además del ingreso $X_1$ hay otras variables que afectan también el consumo, como el número de hijos por familia $X_2$, el sexo $X_3$, la religión $X_4$, la educación $X_5$ y la región geográfica $X_6$. Pero es muy posible que la influencia conjunta de todas o algunas de estas variables sea muy pequeña, o a lo mejor no sistemática ni aleatoria, y que desde el punto de vista práctico y por consideraciones de costo no se justifique su introducción explícita en el modelo. Cabría esperar que su efecto combinado pueda tratarse como una variable aleatoria $u_i$.[10]

4. *Aleatoriedad intrínseca en el comportamiento humano:* Aunque se logre introducir en el modelo todas las variables pertinentes, es posible que se presente alguna aleatoriedad "intrínseca" en $Y$ que no se explique, a pesar de todos los esfuerzos que se inviertan. Las perturbaciones, $u$, pueden reflejar muy bien esta aleatoriedad intrínseca.

5. *Variables representantes (proxy) inadecuadas:* A pesar de que el modelo clásico de regresión (que veremos en el capítulo 3) supone que las variables $Y$ y $X$ se miden con precisión, en la práctica, los datos pueden estar plagados de errores de medición. Consideremos, por ejemplo,

[9] En efecto, en el método de mínimos cuadrados que veremos en el capítulo 3 se supone explícitamente que $E(u_i \mid X_i) = 0$. Ver la sección 3.2.

[10] Otra dificultad es que no es fácil cuantificar variables como sexo, educación y religión.

la famosa teoría de la función de consumo postulada por Milton Friedman.[11] Él considera el *consumo permanente* ($Y^p$) como función del *ingreso permanente* ($X^p$). Pero como la información sobre estas variables no es observable directamente, en la práctica se utilizan variables representantes *(proxy)*, como el consumo actual ($Y$) y el ingreso actual ($X$), que sí son observables. Como las $Y$ y las $X$ observadas pueden no ser iguales a $Y^p$ y $X^p$, existe el problema de errores de medición. El término de perturbación $u$ en este caso puede representar también los errores de medición. Como veremos en un capítulo posterior, de existir tales errores de medición, pueden tener graves repercusiones en la estimación de los coeficientes de regresión, $\beta$.

6. *Principio de parsimonia:* De acuerdo con el principio de la navaja de Occam,[12] conviene mantener el modelo de regresión lo más sencillo posible. Si se explica "sustancialmente" el comportamiento de $Y$ con dos o tres variables explicativas, y si la teoría no es bastante fuerte para indicar otras variables que pueden incluirse, ¿para qué introducir más variables? Que $u_i$ represente todas las demás variables. Por supuesto, no se deben excluir variables pertinentes e importantes sólo para que el modelo de regresión no se complique.

7. *Forma funcional incorrecta:* Aunque se cuente con variables teóricamente correctas para explicar un fenómeno y se obtengan datos sobre ellas, con frecuencia no se conoce la forma de la relación funcional entre la variable regresada y las regresoras. ¿Es el consumo una función lineal (invariable) del ingreso, o es una función no lineal (invariable)? Si se trata de lo primero, $Y_i = \beta_1 + \beta_2 X_1 + u_i$ es la relación funcional adecuada entre $Y$ y $X$, pero en el segundo caso, $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$ puede ser la forma funcional correcta. En los modelos con dos variables, la forma funcional de la relación a menudo se puede inferir del diagrama de dispersión. Sin embargo, en un modelo de regresión múltiple no es fácil determinar la forma funcional apropiada, pues los diagramas de dispersión no se visualizan gráficamente en múltiples dimensiones.

Por todas estas razones, las perturbaciones estocásticas $u_i$ asumen un papel muy valioso en el análisis de regresión, que apreciaremos a medida que avancemos.

## 2.6 Función de regresión muestral (FRM)

Hasta el momento, nos hemos limitado a la población de valores $Y$ que corresponden a valores fijos de $X$. Con toda deliberación evitamos consideraciones muestrales (observe que los datos de la tabla 2.1 representan la población, no una muestra). No obstante, es momento de enfrentar los problemas muestrales, porque en la práctica lo que se tiene al alcance no es más que una muestra de valores de $Y$ que corresponden a algunos valores fijos de $X$. Por tanto, la labor ahora es estimar la FRP con base en información muestral.

A manera de ilustración, supongamos que no se conocía la población de la tabla 2.1 y que la única información que se tenía era una muestra de valores de $Y$ seleccionada al azar para valores dados de $X$ como se presentan en la tabla 2.4. A diferencia de la tabla 2.1, ahora se tiene sólo un valor de $Y$ correspondiente a los valores dados de $X$; cada $Y$ (dada $X_i$) en la tabla 2.4 se selecciona aleatoriamente de las $Y$ similares que corresponden a la misma $X_i$ de la población de la tabla 2.1.

[11] Milton Friedman, *A Theory of the Consumption Function*, Princeton University Press, Princeton, Nueva Jersey, 1957.

[12] "That descriptions be kept as simple as possible until proved inadequate", *The World of Mathematics*, vol. 2, J.R. Newman (ed), Simon & Schuster, Nueva York, 1956, p. 1247, o "Entities should not be multiplied beyond necessity", Donald F. Morrison, *Applied Linear Statistical Methods*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1983, p. 58.

La cuestión es la siguiente: con base en la muestra de la tabla 2.4, ¿es posible predecir el consumo semanal promedio *Y* de la población en su conjunto correspondiente a los valores de *X* seleccionados? En otras palabras, ¿se puede estimar la FRP a partir de los datos de la muestra? Como el lector con seguridad sospecha, quizá no pueda calcularse la FRP con "precisión" debido a las fluctuaciones muestrales. Para ver esto, supongamos que se toma otra muestra aleatoria de la población de la tabla 2.1, la cual se presenta en la tabla 2.5.

Al graficar los datos de las tablas 2.4 y 2.5 se obtiene el diagrama de dispersión de la figura 2.4. En el diagrama de dispersión se trazaron dos líneas de regresión muestral con el fin de "ajustar" razonablemente bien las dispersiones: $FRM_1$ se basa en la primera muestra y $FRM_2$ en la segunda. ¿Cuál de las dos líneas de regresión representa a la línea de regresión poblacional "verdadera"? Si evitamos la tentación de mirar la figura 2.1, que a propósito representa la RP, no hay forma de estar por completo seguro de que alguna de las líneas de regresión de la figura 2.4 representa la verdadera recta (o curva) de regresión poblacional. Las líneas de regresión en

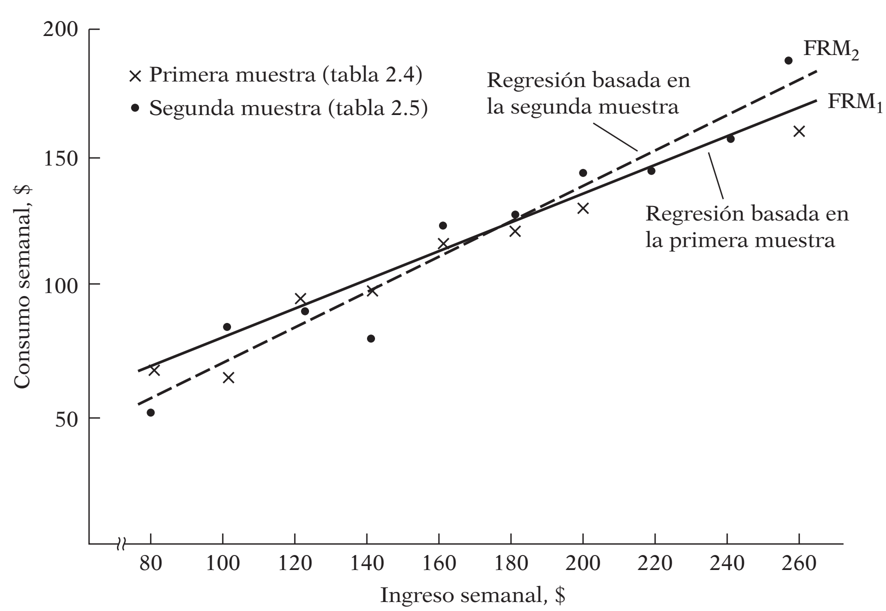
**TABLA 2.4**
**Muestra aleatoria de la población de la tabla 2.1**

| *Y* | *X* |
|---|---|
| 70 | 80 |
| 65 | 100 |
| 90 | 120 |
| 95 | 140 |
| 110 | 160 |
| 115 | 180 |
| 120 | 200 |
| 140 | 220 |
| 155 | 240 |
| 150 | 260 |

**TABLA 2.5**
**Otra muestra aleatoria de la población de la tabla 2.1**

| *Y* | *X* |
|---|---|
| 55 | 80 |
| 88 | 100 |
| 90 | 120 |
| 80 | 140 |
| 118 | 160 |
| 120 | 180 |
| 145 | 200 |
| 135 | 220 |
| 145 | 240 |
| 175 | 260 |

**FIGURA 2.4**
Líneas de regresión basadas en dos muestras diferentes.



la figura 2.4 se conocen como **líneas de regresión muestral**. Se supone que representan la línea de regresión poblacional, pero, debido a fluctuaciones muestrales, son, en el mejor de los casos, sólo una aproximación de la verdadera RP. En general, se obtendrían $N$ FRM diferentes para $N$ muestras diferentes, y estas FRM no por fuerza son iguales.

Ahora, igual que la FRP en la cual se basa la línea de regresión poblacional, se desarrolla el concepto de **función de regresión muestral** (FRM) para representar la línea de regresión muestral. La contraparte muestral de la ecuación (2.2.2) puede escribirse como

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{2.6.1}$$

donde $\hat{Y}$ se lee "$Y$ sombrero" o "$Y$ gorra"

$\hat{Y}_i$ = estimador de $E(Y \mid X_i)$
$\hat{\beta}_1$ = estimador de $\beta_1$
$\hat{\beta}_2$ = estimador de $\beta_2$

Advierta que un **estimador**, conocido también como **estadístico** (muestral), no es más que una regla, fórmula o método para estimar el parámetro poblacional a partir de la información suministrada por la muestra disponible. Un valor numérico particular obtenido por el estimador en un análisis se conoce como **estimación**.[13] Cabe señalar que un estimador es aleatorio, pero una estimación no. (¿Por qué?)

Ahora, tal como la FRP se expresa en dos formas equivalentes, (2.2.2) y (2.4.2), la FRM (2.6.1) se expresa en su forma estocástica de la siguiente manera:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{2.6.2}$$

donde, además de los símbolos ya definidos, $\hat{u}_i$ denota el término **residual** (muestral). Conceptualmente, $\hat{u}_i$ es análogo a $u_i$ y se considera una *estimación* de $u_i$, que se introduce en la FRM por las mismas razones que se introdujo $u_i$ en la FRP.

Así, para resumir, concluimos que el objetivo principal del análisis de regresión es estimar la FRP

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.4.2}$$

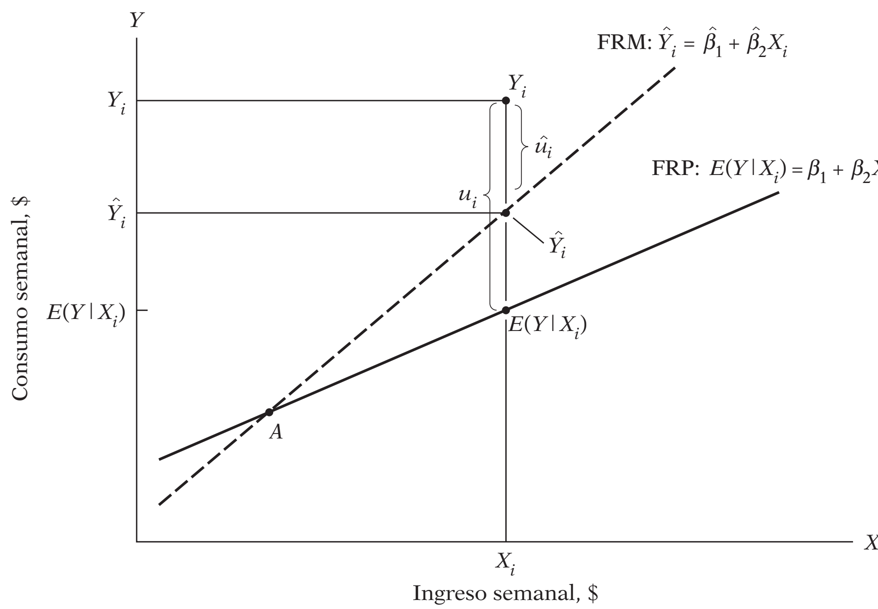
con base en la FRM

$$Y_i = \hat{\beta}_1 + \hat{\beta} x_i + \hat{u}_i \tag{2.6.2}$$

porque son más frecuentes los casos en que el análisis se basa en una sola muestra tomada de una población. Pero, debido a fluctuaciones muestrales, la estimación de la FRP basada en la FRM es, en el mejor de los casos, una aproximación. En la figura 2.5 se presenta gráficamente esta aproximación.

[13] Como mencionamos en la Introducción, un sombrero sobre una variable significa un estimador del valor poblacional pertinente.

**FIGURA 2.5**
Líneas de regresión muestral y poblacional.




Para $X = Xi$ se tiene una observación (muestral), $Y = Y_i$. En términos de la FRM, la $Y_i$ observada se expresa como

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{2.6.3}$$

y en términos de la FRP, como

$$Y_i = E(Y \mid X_i) = u_i \tag{2.6.4}$$

Ahora, como es evidente, en la figura 2.5, $\hat{Y}_i$ *sobrestima* la verdadera $E(Y \mid X_i)$ para la $X_i$ que ahí se muestra. De la misma manera, para cualquier $X_i$, a la izquierda del punto $A$, la FRM *subestimará* la verdadera FRP. Pero el lector puede concluir fácilmente que tal sobrestimación y subestimación es inevitable debido a las fluctuaciones muestrales.

La pregunta crítica es ahora: como la FRM es apenas una aproximación de la FRP, ¿se puede diseñar una regla o método que "acerque" esta aproximación lo más posible? En otras palabras, ¿cómo se debe construir la FRM para que $\hat{\beta}_1$ y $\hat{\beta}_2$ estén tan "cerca" de los verdaderos $\beta_1$ y $\beta_2$ como sea posible, aunque nunca se lleguen a conocer los verdaderos $\beta_1$ y $\beta_2$?

La respuesta ocupará gran parte del capítulo 3. Se advierte aquí que es posible formular procedimientos para que la FRM refleje la FRP tan fielmente como sea posible. Es fascinante considerar esto aunque en realidad nunca se llegue a determinar la propia FRP.

## 2.7 Ejemplos ilustrativos

Este capítulo concluye con dos ejemplos.

**EJEMPLO 2.1**
*Salario promedio por hora según la escolaridad*

La tabla 2.6 proporciona datos sobre el nivel de estudios (medido en años de escolaridad), el salario promedio por hora devengado por las personas por nivel de escolaridad y el número de personas en un nivel de estudios. Ernst Berndt obtuvo originalmente los datos de la tabla de la encuesta de población de mayo de 1985.[14]

Al trazar el salario promedio (condicional) en función del grado de escolaridad, se obtiene la gráfica de la figura 2.6. La curva de regresión de la gráfica muestra la variación de los salarios promedio de acuerdo con el grado de escolaridad; por lo general, aquéllos se incrementan a la par que el grado de escolaridad, conclusión que no debe causar sorpresa. En un capítulo posterior examinaremos la influencia de otras variables en el salario promedio.

**TABLA 2.6**
**Salario promedio por hora según la escolaridad**

Fuente: Arthur S. Goldberger, *Introductory Econometrics,* Harvard University Press, Cambridge, Massachusetts, 1998, tabla 1.1, p. 5 (adaptada).

| Años de escolaridad | Salario promedio, $ | | Número de personas |
|---|---|---|---|
| 6 | 4.4567 | | 3 |
| 7 | 5.7700 | | 5 |
| 8 | 5.9787 | | 15 |
| 9 | 7.3317 | | 12 |
| 10 | 7.3182 | | 17 |
| 11 | 6.5844 | | 27 |
| 12 | 7.8182 | | 218 |
| 13 | 7.8351 | | 37 |
| 14 | 11.0223 | | 56 |
| 15 | 10.6738 | | 13 |
| 16 | 10.8361 | | 70 |
| 17 | 13.6150 | | 24 |
| 18 | 13.5310 | | 31 |
| | | Total | 528 |

**FIGURA 2.6**
Relación entre salario promedio y escolaridad.

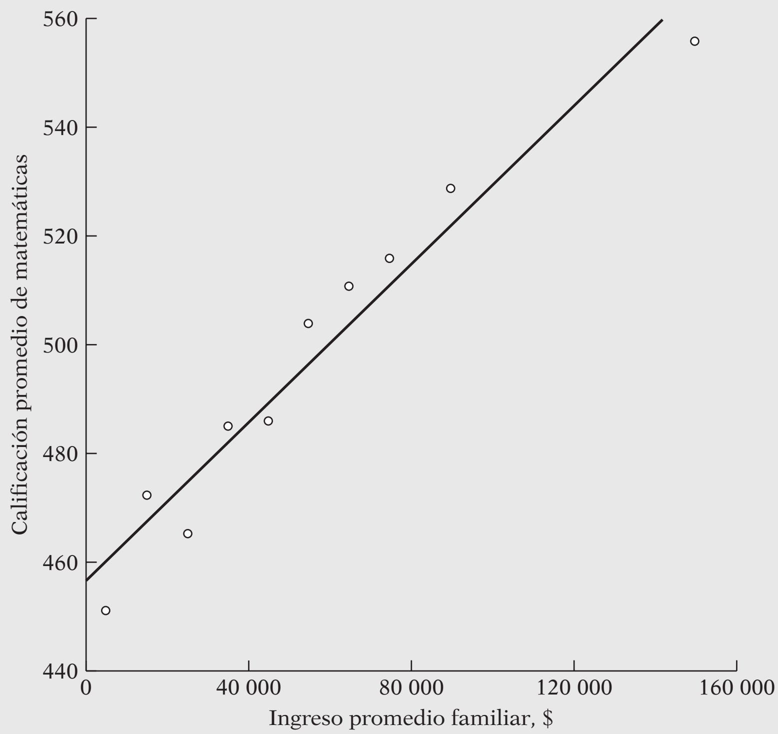
[14] Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison Wesley, Reading, Massachusetts, 1991. Por cierto, es un excelente libro que el lector tal vez quiera leer para averiguar cómo investigan los econometristas.

**EJEMPLO 2.2**
*Calificaciones de matemáticas en el examen SAT según ingreso familiar*

En la tabla 2.10 del ejercicio 2.17 hay datos sobre el promedio de calificaciones del examen de aptitud académica (SAT, *Scholastic Aptitude Test*) de lectura crítica, matemáticas y redacción de 947 347 estudiantes que solicitaron su admisión a licenciatura en 2007. Se representó gráficamente la calificación promedio de matemáticas en función del ingreso familiar promedio para obtener la ilustración de la figura 2.7.

*Nota:* En virtud de los intervalos abiertos de ingreso para la primera y última categorías de la tabla 2.10, se supone que el ingreso promedio familiar más bajo es de 5 000 dólares, y el más alto, de 150 000 dólares.

**FIGURA 2.7**
Relación entre la calificación promedio de matemáticas en el examen SAT y el ingreso familiar promedio.




Como muestra la figura 2.7, la calificación promedio de matemáticas aumenta conforme lo hace el ingreso familiar. Como el número de estudiantes que presentaron el examen SAT es muy grande, es probable que represente a toda la población de estudiantes que presentaron el examen. Por tanto, también es probable que la línea de regresión trazada en la figura 2.7 represente la línea de regresión poblacional.

Puede haber varias razones para la relación positiva entre ambas variables. Por ejemplo, se puede argumentar que los estudiantes con un mayor ingreso familiar están en mejores condiciones de costear clases privadas para el examen. Además, es más probable que los estudiantes de mayores ingresos tengan padres con educación superior. También es posible que los estudiantes con calificaciones más altas de matemáticas procedan de mejores escuelas. A usted se le pueden ocurrir otras explicaciones de la relación positiva entre las dos variables.

## Resumen y conclusiones

1. El concepto fundamental del análisis de regresión es el de **función de esperanza condicional (FEC)**, o **función de regresión poblacional (FRP)**. El objetivo del análisis de regresión es averiguar la forma en que varía el valor promedio de la variable dependiente (o regresada) de acuerdo con el valor dado de la variable explicativa (o regresora).
2. Este libro trata sobre todo de **FRP lineales**, es decir, regresiones lineales en los parámetros. Éstas pueden ser o no lineales en la variable regresada o las regresoras.
3. Para efectos prácticos, la **FRP estocástica** es la que importa. El **término de perturbación estocástica** $u_i$ desempeña una función crucial para estimar la FRP.
4. La FRP es un concepto idealizado, pues en la práctica pocas veces se tiene acceso al total de la población de interés. Por lo general se cuenta sólo con una muestra de observaciones de la población. En consecuencia, se utiliza la **función de regresión muestral estocástica (FRM)** para estimar la FRP; la forma de lograrlo se analiza en el capítulo 3.

## EJERCICIOS

### Preguntas

2.1. ¿Cuál es la función de esperanza condicional o función de regresión poblacional?

2.2. ¿Cuál es la diferencia entre la función de regresión poblacional y la función de regresión muestral? ¿Se trata de distintos nombres para la misma función?

2.3. ¿Qué papel desempeña el término de error estocástico $u_i$ en el análisis de regresión? ¿Cuál es la diferencia entre el término de error estocástico y el residual $\hat{u}_i$?

2.4. ¿Por qué es necesario el análisis de regresión? ¿Por qué no tan sólo utilizar el valor medio de la variable regresada como su mejor valor?

2.5. ¿Qué se quiere dar a entender con modelo de regresión *lineal*?

2.6. Determine si los siguientes modelos son lineales en los parámetros, en las variables o en ambos. ¿Cuáles de estos modelos son de regresión lineal?

| Modelo | Título descriptivo |
|---|---|
| *a*) $Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$ | Recíproco |
| *b*) $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$ | Semilogarítmico |
| *c*) $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$ | Semilogarítmico inverso |
| *d*) $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$ | Logarítmico o doble logarítmico |
| *e*) $\ln Y_i = \beta_1 - \beta_2 \left(\frac{1}{X_i}\right) + u_i$ | Logarítmico recíproco |

*Nota:* ln = logaritmo natural (es decir, logaritmo base e); $u_i$ es el término de perturbación estocástica. Estudiaremos estos modelos en el capítulo 6.

2.7. ¿Son modelos de regresión lineal los siguientes? ¿Por qué?

*a*) $Y_i = e^{\beta_1 + \beta_2 X_i + u_i}$

*b*) $Y_i = \dfrac{1}{1 + e^{\beta_1 + \beta_2 X_i + u_i}}$

*c*) $\ln Y_i = \beta_1 + \beta_2 \left(\dfrac{1}{X_i}\right) + u_i$

*d*) $Y_i = \beta_1 + (0.75 - \beta_1)e^{-\beta_2(X_i - 2)} + u_i$

*e*) $Y_i = \beta_1 + \beta_2^3 X_i + u_i$

2.8. ¿Qué se entiende por un modelo de regresión *intrínsecamente lineal*? Si en el ejercicio 2.7*d*) $\beta_2$ valiera 0.8, ¿sería un modelo de regresión lineal o no lineal?

2.9. Considere los siguientes modelos no estocásticos (es decir, modelos sin el término de error estocástico). ¿Son lineales estos modelos de regresión? De no serlo, ¿sería posible, con manipulaciones algebraicas apropiadas, convertirlos en modelos lineales?

*a*) $Y_i = \dfrac{1}{\beta_1 + \beta_2 X_i}$

*b*) $Y_i = \dfrac{X_i}{\beta_1 + \beta_2 X_i}$

*c*) $Y_i = \dfrac{1}{1 + \exp(-\beta_1 - \beta_2 X_i)}$

2.10. Considere el diagrama de dispersión de la figura 2.8 junto con la línea de regresión. ¿Qué conclusión general deduce de este diagrama? ¿La línea de regresión del diagrama es una línea de regresión poblacional o una línea de regresión muestral?

2.11. Del diagrama de dispersión de la figura 2.9, ¿qué conclusiones generales deduce? ¿En qué teoría económica se basa este diagrama de dispersión? (*Pista*: busque cualquier libro de texto de economía internacional y estudie el modelo de comercio Heckscher-Ohlin).

2.12. ¿Qué revela el diagrama de dispersión de la figura 2.10? Con base en dicho diagrama, ¿se puede decir que las leyes del salario mínimo propician el bienestar económico?

2.13. ¿La línea de regresión de la figura I.3, en la Introducción, es la FRP o la FRM? ¿Por qué? ¿Cómo se interpretarían los puntos alrededor de la línea de regresión? Además del PIB, ¿qué otros factores, o variables, determinarían el consumo personal?

**FIGURA 2.8**

Tasas de crecimiento de los salarios y exportaciones reales del sector manufacturero. Los datos corresponden a 50 países en vías de desarrollo durante 1970-1990.

Fuente: Banco Mundial, *World Development Report 1995,* p. 55. Las fuentes originales son datos de ONUDI y del Banco Mundial.

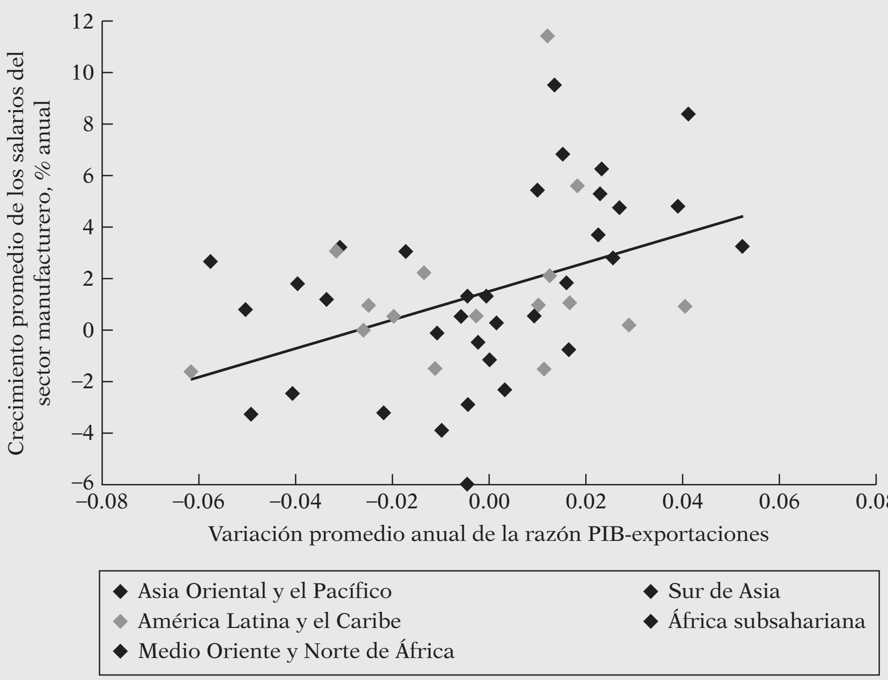


**FIGURA 2.9**
Intensidad de la capacidad en las exportaciones y dotación de capital humano. Los datos corresponden a 126 países industrializados y en vías de desarrollo en 1985. Los valores trazados en el eje horizontal son logaritmos de la razón entre la escolaridad promedio alcanzada en cada país y la superficie de tierra cultivable del país; los valores del eje vertical son logaritmos de la razón entre las exportaciones de productos manufacturados y productos básicos.

Fuente: Banco Mundial, *World Development Report 1995,* p. 59. Fuentes originales: datos de exportación de la Oficina de Estadística de las Naciones Unidas, base de datos COMTRADE; datos sobre educación del Programa de las Naciones Unidas para el Desarrollo (PNUD), 1990; datos sobre el uso de la tierra del Banco Mundial.

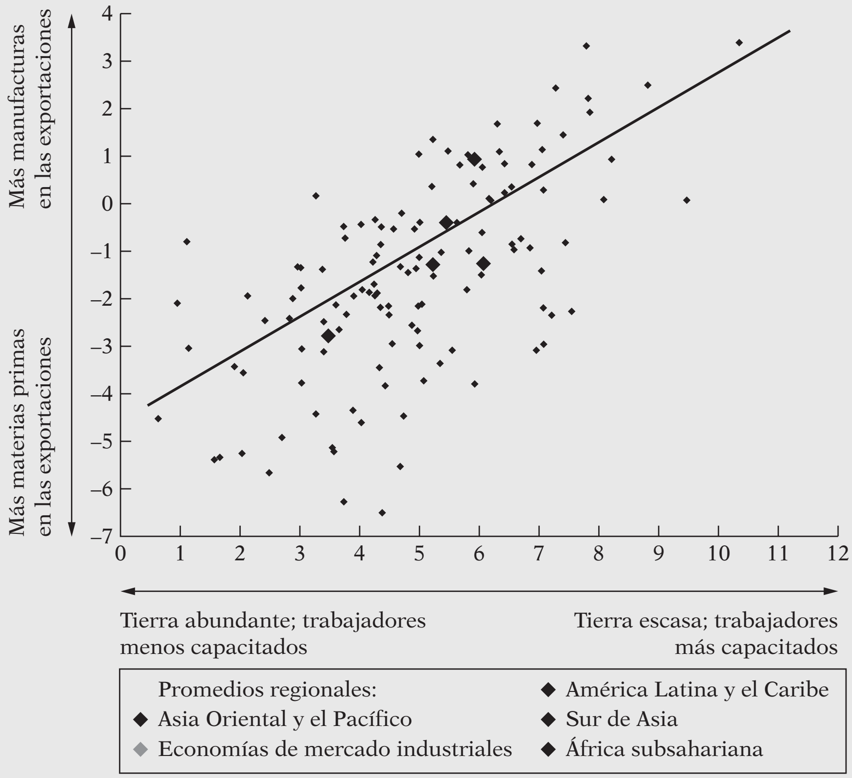
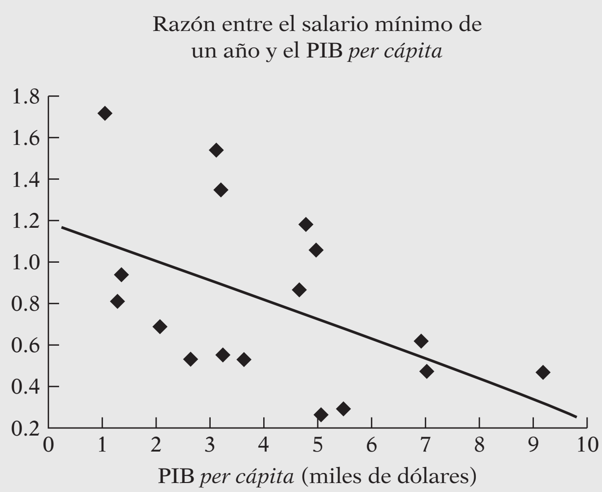



**FIGURA 2.10**
Salario mínimo y PIB *per cápita*. La muestra consta de 17 países en vías de desarrollo. Los años varían por país de 1988 a 1992. Los datos se dan en precios internacionales.

Fuente: Banco Mundial, *World Development Report 1995,* p. 75.




### Ejercicios empíricos

2.14. Se proporcionan los datos de la tabla 2.7 correspondientes a Estados Unidos de 1980 a 2006.

*a*) Grafique la tasa de participación de la fuerza laboral civil masculina en función de la tasa de desempleo civil para los hombres. Trace a mano una línea de regresión a través

**TABLA 2.7**
**Datos sobre la participación de la fuerza laboral de Estados Unidos en 1980-2006**

Fuente: *Economic Report of the President, 2007.*

| Año | TPFLCM[1] | TPFLCF[2] | TDCH[3] | TDCM[4] | IPH82[5] | IPH[6] |
|---|---|---|---|---|---|---|
| 1980 | 77.40000 | 51.50000 | 6.900000 | 7.400000 | 7.990000 | 6.840000 |
| 1981 | 77.00000 | 52.10000 | 7.400000 | 7.900000 | 7.880000 | 7.430000 |
| 1982 | 76.60000 | 52.60000 | 9.900000 | 9.400000 | 7.860000 | 7.860000 |
| 1983 | 76.40000 | 52.90000 | 9.900000 | 9.200000 | 7.950000 | 8.190000 |
| 1984 | 76.40000 | 53.60000 | 7.400000 | 7.600000 | 7.950000 | 8.480000 |
| 1985 | 76.30000 | 54.50000 | 7.000000 | 7.400000 | 7.910000 | 8.730000 |
| 1986 | 76.30000 | 55.30000 | 6.900000 | 7.100000 | 7.960000 | 8.920000 |
| 1987 | 76.20000 | 56.00000 | 6.200000 | 6.200000 | 7.860000 | 9.130000 |
| 1988 | 76.20000 | 56.60000 | 5.500000 | 5.600000 | 7.810000 | 9.430000 |
| 1989 | 76.40000 | 57.40000 | 5.200000 | 5.400000 | 7.750000 | 9.800000 |
| 1990 | 76.40000 | 57.50000 | 5.700000 | 5.500000 | 7.660000 | 10.190000 |
| 1991 | 75.80000 | 57.40000 | 7.200000 | 6.400000 | 7.580000 | 10.500000 |
| 1992 | 75.80000 | 57.80000 | 7.900000 | 7.000000 | 7.550000 | 10.760000 |
| 1993 | 75.40000 | 57.90000 | 7.200000 | 6.600000 | 7.520000 | 11.030000 |
| 1994 | 75.10000 | 58.80000 | 6.200000 | 6.000000 | 7.530000 | 11.320000 |
| 1995 | 75.00000 | 58.90000 | 5.600000 | 5.600000 | 7.530000 | 11.640000 |
| 1996 | 74.90000 | 59.30000 | 5.400000 | 5.400000 | 7.570000 | 12.030000 |
| 1997 | 75.00000 | 59.80000 | 4.900000 | 5.000000 | 7.680000 | 12.490000 |
| 1998 | 74.90000 | 59.80000 | 4.400000 | 4.600000 | 7.890000 | 13.000000 |
| 1999 | 74.70000 | 60.00000 | 4.100000 | 4.300000 | 8.000000 | 13.470000 |
| 2000 | 74.80000 | 59.90000 | 3.900000 | 4.100000 | 8.030000 | 14.000000 |
| 2001 | 74.40000 | 59.80000 | 4.800000 | 4.700000 | 8.110000 | 14.530000 |
| 2002 | 74.10000 | 59.60000 | 5.900000 | 5.600000 | 8.240000 | 14.950000 |
| 2003 | 73.50000 | 59.50000 | 6.300000 | 5.700000 | 8.270000 | 15.350000 |
| 2004 | 73.30000 | 59.20000 | 5.600000 | 5.400000 | 8.230000 | 15.670000 |
| 2005 | 73.30000 | 59.30000 | 5.100000 | 5.100000 | 8.170000 | 16.110000 |
| 2006 | 73.50000 | 59.40000 | 4.600000 | 4.600000 | 8.230000 | 16.730000 |

Las siguientes menciones se refieren al documento original.
[1] TPFLCM = Tasa de participación de la fuerza laboral civil masculina (%), tabla B-39, p. 277.
[2] TPFLCF = Tasa de participación de la fuerza laboral civil femenina (%), tabla B-39, p. 277.
[3] TDCH = Tasa de desempleo civil, hombres (%), tabla B-42, p. 280.
[4] TDCM = Tasa de desempleo civil, mujeres (%), tabla B-42, p. 280.
[5] IPH82 = Ingresos promedio por hora (dólares de 1982), tabla B-47, p. 286.
[6] IPH = Ingresos promedio por hora (dólares actuales), tabla B-47, p. 286.

de los puntos de dispersión. Mencione *a priori* la relación esperada entre ambas tasas y comente cuál es la teoría económica que sirve de fundamento. ¿Este diagrama de dispersión apoya dicha teoría?

*b*) Repita el inciso *a*) para las mujeres.

*c*) Ahora grafique las tasas de participación laboral de ambos sexos en función de los ingresos promedio por hora (en dólares de 1982). (Quizá convenga utilizar diagramas independientes.) Ahora, ¿qué concluye? ¿Cómo racionalizaría esa conclusión?

*d*) ¿Se puede trazar la tasa de participación de la fuerza laboral en función de la tasa de desempleo y de los ingresos promedio por hora, de manera simultánea? Si no fuera así, ¿cómo expresaría verbalmente la relación entre esas tres variables?

2.15. En la tabla 2.8 se proporcionan los datos sobre gasto en comida y gasto total (en rupias) para una muestra de 55 familias rurales de India. (A principios de 2000, un dólar estadounidense equivalía a casi 40 rupias indias.)

*a*) Grafique los datos con el eje vertical para el gasto en comida y el eje horizontal para el gasto total; trace una línea de regresión a través de los puntos de dispersión.

*b*) ¿Qué conclusiones generales se pueden deducir de este ejemplo?

**TABLA 2.8** **Gasto total y en comida (rupias)**

| Observación | Gasto en comida | Gasto total | Observación | Gasto en comida | Gasto total |
|---|---|---|---|---|---|
| 1 | 217.0000 | 382.0000 | 29 | 390.0000 | 655.0000 |
| 2 | 196.0000 | 388.0000 | 30 | 385.0000 | 662.0000 |
| 3 | 303.0000 | 391.0000 | 31 | 470.0000 | 663.0000 |
| 4 | 270.0000 | 415.0000 | 32 | 322.0000 | 677.0000 |
| 5 | 325.0000 | 456.0000 | 33 | 540.0000 | 680.0000 |
| 6 | 260.0000 | 460.0000 | 34 | 433.0000 | 690.0000 |
| 7 | 300.0000 | 472.0000 | 35 | 295.0000 | 695.0000 |
| 8 | 325.0000 | 478.0000 | 36 | 340.0000 | 695.0000 |
| 9 | 336.0000 | 494.0000 | 37 | 500.0000 | 695.0000 |
| 10 | 345.0000 | 516.0000 | 38 | 450.0000 | 720.0000 |
| 11 | 325.0000 | 525.0000 | 39 | 415.0000 | 721.0000 |
| 12 | 362.0000 | 554.0000 | 40 | 540.0000 | 730.0000 |
| 13 | 315.0000 | 575.0000 | 41 | 360.0000 | 731.0000 |
| 14 | 355.0000 | 579.0000 | 42 | 450.0000 | 733.0000 |
| 15 | 325.0000 | 585.0000 | 43 | 395.0000 | 745.0000 |
| 16 | 370.0000 | 586.0000 | 44 | 430.0000 | 751.0000 |
| 17 | 390.0000 | 590.0000 | 45 | 332.0000 | 752.0000 |
| 18 | 420.0000 | 608.0000 | 46 | 397.0000 | 752.0000 |
| 19 | 410.0000 | 610.0000 | 47 | 446.0000 | 769.0000 |
| 20 | 383.0000 | 616.0000 | 48 | 480.0000 | 773.0000 |
| 21 | 315.0000 | 618.0000 | 49 | 352.0000 | 773.0000 |
| 22 | 267.0000 | 623.0000 | 50 | 410.0000 | 775.0000 |
| 23 | 420.0000 | 627.0000 | 51 | 380.0000 | 785.0000 |
| 24 | 300.0000 | 630.0000 | 52 | 610.0000 | 788.0000 |
| 25 | 410.0000 | 635.0000 | 53 | 530.0000 | 790.0000 |
| 26 | 220.0000 | 640.0000 | 54 | 360.0000 | 795.0000 |
| 27 | 403.0000 | 648.0000 | 55 | 305.0000 | 801.0000 |
| 28 | 350.0000 | 650.0000 | | | |

Fuente: Chandan Mukherjee, Howard White y Marc Wuyts, *Econometrics and Data Analysis for Developing Countries,* Routledge, Nueva York, 1998, p. 457.

*c*) Diga *a priori* si se esperaría que el gasto en comida se incrementara de manera lineal conforme el gasto total aumentase, independientemente del nivel de gasto. ¿Por qué? Puede emplear el gasto total como representante del ingreso total.

2.16. La tabla 2.9 presenta datos sobre el promedio de calificaciones del examen de aptitud académica SAT de los estudiantes que solicitaron admisión a licenciatura de 1972 a 2007. Estos datos representan las calificaciones en el examen de lectura crítica y matemáticas de hombres y mujeres. La categoría de redacción se introdujo en 2006. Por tanto, estos datos no se incluyen.

*a*) Con el eje horizontal para los años y el vertical para las calificaciones del examen SAT, grafique las calificaciones de lectura crítica y matemáticas de hombres y mujeres por separado.

*b*) ¿Qué conclusiones generales se obtienen?

*c*) Al conocer las calificaciones de lectura crítica de hombres y mujeres, ¿cómo haría para predecir las calificaciones de matemáticas?

*d*) Grafique las calificaciones de matemáticas de las mujeres contra las calificaciones de matemáticas de los hombres. ¿Qué observa?

**TABLA 2.9**
**Promedio grupal total de las calificaciones del examen de razonamiento SAT: estudiantes que solicitaron ingreso a licenciatura, 1972-2007**

Fuente: College Board, 2007.

| | Lectura crítica | | | Matemáticas | | |
|---|---|---|---|---|---|---|
| Año | Hombres | Mujeres | Total | Hombres | Mujeres | Total |
| 1972 | 531 | 529 | 530 | 527 | 489 | 509 |
| 1973 | 523 | 521 | 523 | 525 | 489 | 506 |
| 1974 | 524 | 520 | 521 | 524 | 488 | 505 |
| 1975 | 515 | 509 | 512 | 518 | 479 | 498 |
| 1976 | 511 | 508 | 509 | 520 | 475 | 497 |
| 1977 | 509 | 505 | 507 | 520 | 474 | 496 |
| 1978 | 511 | 503 | 507 | 517 | 474 | 494 |
| 1979 | 509 | 501 | 505 | 516 | 473 | 493 |
| 1980 | 506 | 498 | 502 | 515 | 473 | 492 |
| 1981 | 508 | 496 | 502 | 516 | 473 | 492 |
| 1982 | 509 | 499 | 504 | 516 | 473 | 493 |
| 1983 | 508 | 498 | 503 | 516 | 474 | 494 |
| 1984 | 511 | 498 | 504 | 518 | 478 | 497 |
| 1985 | 514 | 503 | 509 | 522 | 480 | 500 |
| 1986 | 515 | 504 | 509 | 523 | 479 | 500 |
| 1987 | 512 | 502 | 507 | 523 | 481 | 501 |
| 1988 | 512 | 499 | 505 | 521 | 483 | 501 |
| 1989 | 510 | 498 | 504 | 523 | 482 | 502 |
| 1990 | 505 | 496 | 500 | 521 | 483 | 501 |
| 1991 | 503 | 495 | 499 | 520 | 482 | 500 |
| 1992 | 504 | 496 | 500 | 521 | 484 | 501 |
| 1993 | 504 | 497 | 500 | 524 | 484 | 503 |
| 1994 | 501 | 497 | 499 | 523 | 487 | 504 |
| 1995 | 505 | 502 | 504 | 525 | 490 | 506 |
| 1996 | 507 | 503 | 505 | 527 | 492 | 508 |
| 1997 | 507 | 503 | 505 | 530 | 494 | 511 |
| 1998 | 509 | 502 | 505 | 531 | 496 | 512 |
| 1999 | 509 | 502 | 505 | 531 | 495 | 511 |
| 2000 | 507 | 504 | 505 | 533 | 498 | 514 |
| 2001 | 509 | 502 | 506 | 533 | 498 | 514 |
| 2002 | 507 | 502 | 504 | 534 | 500 | 516 |
| 2003 | 512 | 503 | 507 | 537 | 503 | 519 |
| 2004 | 512 | 504 | 508 | 537 | 501 | 518 |
| 2005 | 513 | 505 | 508 | 538 | 504 | 520 |
| 2006 | 505 | 502 | 503 | 536 | 502 | 518 |
| 2007 | 504 | 502 | 502 | 533 | 499 | 515 |

*Nota:* Para 1972-1986 se aplicó una fórmula a la media y a la desviación estándar originales para convertir la media a la escala recentrada. Para 1987-1995, las calificaciones de cada estudiante se convirtieron a la escala recentrada y luego se volvió a calcular la media. De 1996 a 1999, casi todos los estudiantes recibieron calificaciones según la escala recentrada. Toda calificación basada en la escala original se convirtió a la escala recentrada antes de calcular la media. De 2000 a 2007, todas las calificaciones se basaron en la escala recentrada.

2.17. La tabla 2.10 presenta datos sobre las calificaciones del examen de razonamiento SAT clasificadas por ingreso para tres tipos de pruebas: lectura crítica, matemáticas y redacción. En el ejemplo 2.2 se presentó la figura 2.7, que contiene una gráfica de la media de las calificaciones de matemáticas en función del ingreso familiar promedio.

*a*) Consulte la figura 2.7 y prepare una gráfica parecida que relacione el promedio de calificaciones en lectura crítica con el ingreso familiar promedio. Compare sus resultados con los de la figura 2.7.

**TABLA 2.10**
**Examen de razonamiento SAT clasificado por ingreso familiar**

Fuente: College Board, 2007, Estudiantes que solicitaron ingreso a licenciatura, tabla 11.

| Ingreso familiar ($) | Número de estudiantes que presentaron el examen | Lectura crítica | | Matemáticas | | Redacción | |
|---|---|---|---|---|---|---|---|
| | | Media | DE | Media | DE | Media | DE |
| <10 000 | 40 610 | 427 | 107 | 451 | 122 | 423 | 104 |
| 10 000-20 000 | 72 745 | 453 | 106 | 472 | 113 | 446 | 102 |
| 20 000-30 000 | 61 244 | 454 | 102 | 465 | 107 | 444 | 97 |
| 30 000-40 000 | 83 685 | 476 | 103 | 485 | 106 | 466 | 98 |
| 40 000-50 000 | 75 836 | 489 | 103 | 486 | 105 | 477 | 99 |
| 50 000-60 000 | 80 060 | 497 | 102 | 504 | 104 | 486 | 98 |
| 60 000-70 000 | 75 763 | 504 | 102 | 511 | 103 | 493 | 98 |
| 70 000-80 000 | 81 627 | 508 | 101 | 516 | 103 | 498 | 98 |
| 80 000-100 000 | 130 752 | 520 | 102 | 529 | 104 | 510 | 100 |
| >100 000 | 245 025 | 544 | 105 | 556 | 107 | 537 | 103 |

*b*) Repita el inciso *a*) para relacionar el promedio de calificaciones de redacción con el ingreso familiar promedio, y compare sus resultados con los de las otras dos gráficas.

*c*) Examine las tres gráficas y mencione qué conclusión general puede obtener.

Capítulo 3

# Modelo de regresión con dos variables: problema de estimación

Como explicamos en el capítulo 2, la primera tarea consiste en estimar la función de regresión poblacional (FRP) con base en la función de regresión muestral (FRM) en la forma más precisa posible. En el **apéndice A** se analizan dos métodos de estimación frecuentes: 1) **mínimos cuadrados ordinarios (MCO)** y 2) **máxima verosimilitud (MV)**. El método de MCO es el más común en el análisis de regresión, sobre todo por ser mucho más intuitivo y matemáticamente más sencillo que el método de máxima verosimilitud. Además, como veremos más adelante, en el contexto de la regresión lineal, por lo general los dos métodos proporcionan resultados similares.

## 3.1 Método de mínimos cuadrados ordinarios (MCO)

El método de mínimos cuadrados ordinarios se atribuye a Carl Friedrich Gauss, matemático alemán. A partir de ciertos supuestos (estudiados en la sección 3.2), el método de mínimos cuadrados presenta propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Para entenderlo, primero explicaremos el principio de los mínimos cuadrados.

Recuerde la FRP de dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(2.4.2)}$$

Sin embargo, como mencionamos en el capítulo 2, la FRP no es observable directamente. Se calcula a partir de la FRM:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad \textbf{(2.6.2)}$$

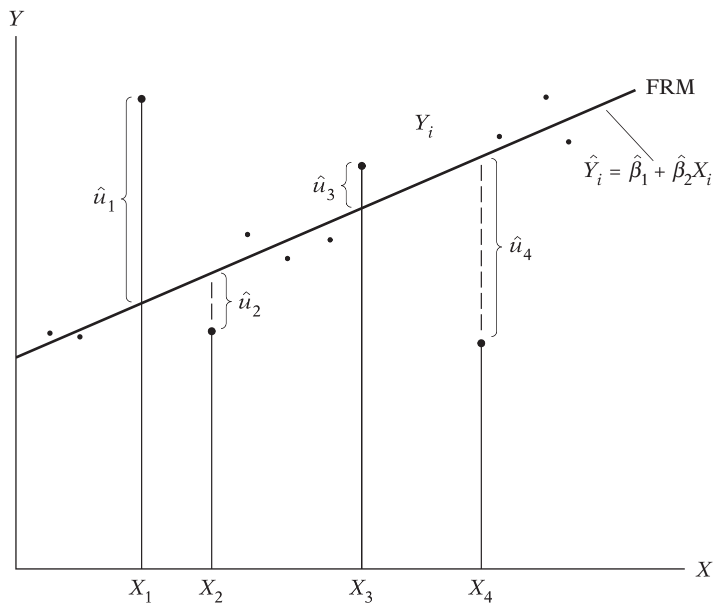
$$= \hat{Y}_i + \hat{u}_i \qquad \textbf{(2.6.3)}$$

donde $\hat{Y}_i$ es el valor estimado (media condicional) de $Y_i$.

Pero, ¿cómo se determina la FRM? Para hacerlo, se procede de la siguiente forma. Primero, se expresa la ecuación (2.6.3) como

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned} \qquad \textbf{(3.1.1)}$$

**FIGURA 3.1**
Criterio de mínimos cuadrados.




que muestra que los $\hat{u}_i$ (los residuos) son simplemente las diferencias entre los valores observados y los estimados de *Y*.

Ahora, dados *n* pares de observaciones de *Y* y *X*, nos interesa determinar la FRM de manera que quede lo más cerca posible de la *Y* observada. Con este fin, se adopta el siguiente criterio: seleccionar la FRM de modo que la suma de los residuos $\sum \hat{u}_i = \sum (Y_i - \hat{Y}_i)$ sea la menor posible. Este criterio, aunque es intuitivamente atractivo, no es muy bueno, como se ve en el diagrama de dispersión hipotético de la figura 3.1.

Si se adopta el criterio de reducir $\sum \hat{u}_i$, la figura 3.1 muestra que los residuos $\hat{u}_2$ y $\hat{u}_3$, al igual que los residuos $\hat{u}_1$ y $\hat{u}_4$, reciben el mismo peso en la suma $(\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4)$, aunque los dos primeros están mucho más cerca la FRM que los dos últimos. En otras palabras, a todos los residuos se les da la misma importancia sin considerar cuán cerca o cuán dispersas estén las observaciones individuales de la FRM. Debido a lo anterior, es muy posible que la suma algebraica de las $\hat{u}_i$ sea pequeña (aun cero) a pesar de que las $\hat{u}_i$ estén muy dispersas respecto de la FRM. Para verificar lo anterior, $\hat{u}_1$, $\hat{u}_2$, $\hat{u}_3$ y $\hat{u}_4$ en la figura 3.1 asumirán respectivamente los valores de 10, −2, +2 y −10. La suma algebraica de estos residuos es cero a pesar de que $\hat{u}_1$ y $\hat{u}_4$ presentan una mayor dispersión alrededor de FRM que $\hat{u}_2$ y $\hat{u}_3$. Se evita este problema con el *criterio de mínimos cuadrados*, el cual establece que la FRM se determina en forma tal que

$$\begin{aligned} \sum \hat{u}_i^2 &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \end{aligned} \tag{3.1.2}$$

sea lo más pequeña posible, donde $\hat{u}_i^2$ son los residuos elevados al cuadrado. Al elevar al cuadrado $\hat{u}_i$, este método da más peso a los residuos como $\hat{u}_1$ y $\hat{u}_4$ en la figura 3.1 que a los residuos $\hat{u}_2$ y $\hat{u}_3$. Como ya anotamos, con el criterio de $\sum \hat{u}_i$ mínima, la suma puede ser pequeña a pesar de que los $\hat{u}_i$ estén muy dispersos alrededor de la FRM. La situación anterior no se presenta con el procedimiento de mínimos cuadrados, pues, entre mayor sea $\hat{u}_i$ (en valor absoluto), mayor será $\sum \hat{u}_i^2$. Otra justificación del método de mínimos cuadrados es que los estimadores obtenidos con este método tienen algunas propiedades estadísticas muy deseables, como veremos en breve.

**TABLA 3.1**
**Determinación experimental de la FRM**

| $Y_i$ (1) | $X_t$ (2) | $\hat{Y}_{1i}$ (3) | $\hat{u}_{1i}$ (4) | $\hat{u}_{1i}^2$ (5) | $\hat{Y}_{2i}$ (6) | $\hat{u}_{2i}$ (7) | $\hat{u}_{2i}^2$ (8) |
|---|---|---|---|---|---|---|---|
| 4 | 1 | 2.929 | 1.071 | 1.147 | 4 | 0 | 0 |
| 5 | 4 | 7.000 | –2.000 | 4.000 | 7 | –2 | 4 |
| 7 | 5 | 8.357 | –1.357 | 1.841 | 8 | –1 | 1 |
| 12 | 6 | 9.714 | 2.286 | 5.226 | 9 | 3 | 9 |
| Suma: 28 | 16 | | 0.0 | 12.214 | | 0 | 14 |

*Notas:* $\hat{Y}_{1i} = 1.572 + 1.357X_i$ (es decir, $\hat{\beta}_1 = 1.572$ y $\hat{\beta}_2 = 1.357$)
$\hat{Y}_{2i} = 3.0 + 1.0X_i$ (es decir, $\hat{\beta}_1 = 3$ y $\hat{\beta}_2 = 1.0$)
$\hat{u}_{1i} = (Y_i - \hat{Y}_{1i})$
$\hat{u}_{2i} = (Y_i - \hat{Y}_{2i})$

A partir de la ecuación (3.1.2) es evidente que

$$\sum \hat{u}_i^2 = f(\hat{\beta}_1, \hat{\beta}_2) \tag{3.1.3}$$

es decir, la suma de los residuos elevados al cuadrado es algún tipo de función de los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$. Por cada conjunto dado de datos con diferentes valores para $\hat{\beta}_1$ y $\hat{\beta}_2$, se obtendrá como resultado $\hat{u}$ diferentes y, por consiguiente, valores diferentes de $\sum \hat{u}_i^2$. Para ver esto claramente, consideremos las cifras hipotéticas de $Y$ y de $X$ de las primeras dos columnas de la tabla 3.1. Realicemos ahora dos experimentos. En el experimento 1, sea $\hat{\beta}_1 = 1.572$ y $\hat{\beta}_2 = 1.357$ (por ahora no preocupa la forma como se obtuvieron estos valores; es decir, se trata sólo de conjeturas).[1] Con estos valores $\hat{\beta}$ y los valores de $X$ dados en la columna (2) de la tabla 3.1, se calcula fácilmente la $Y_i$ estimada dada en la columna (3) de la tabla y denotada $\hat{Y}_{1i}$ (el subíndice 1 indica el primer experimento). Ahora realicemos otro experimento, pero esta vez con los valores de $\hat{\beta}_1 = 3$ y $\hat{\beta}_2 = 1$. Los valores estimados de $Y_i$ a partir de este experimento están dados por $\hat{Y}_{2i}$ en la columna (6) de la tabla 3.1. Como los valores de $\hat{\beta}$ en los dos experimentos son diferentes, se obtienen también valores diferentes para los residuos estimados, como aparece en la tabla; $\hat{u}_{1i}$ corresponden a los del primer experimento y $\hat{u}_{2i}$ corresponden a los del segundo. Los cuadrados de estos residuos están dados en las columnas (5) y (8). Como es lógico, según se esperaba de la ecuación (3.1.3), estas sumas de residuos al cuadrado son diferentes, pues se basan en conjuntos diferentes de valores de $\hat{\beta}$.

Ahora, ¿qué conjuntos de $\hat{\beta}$ se deben escoger? Como los valores de $\hat{\beta}$ del primer experimento dan una $\sum \hat{u}_i^2$ (= 12.214) inferior a la que se obtiene con los valores de $\hat{\beta}$ del segundo experimento (= 14), se puede decir que las $\hat{\beta}$ estimadas del primer experimento son los "mejores" valores. Pero, ¿cómo saberlo?: si se tuviera tiempo y paciencia infinitos, se podrían realizar muchos más experimentos de este tipo, escogiendo cada vez diferentes conjuntos de $\hat{\beta}$ y comparando las $\sum \hat{u}_i^2$ resultantes, y luego escogiendo el conjunto de valores de $\hat{\beta}$ que diera el menor valor posible de $\sum \hat{u}_i^2$, y suponiendo, desde luego, que se consideraron todos los valores posibles de $\beta_1$ y $\beta_2$. Pero como el tiempo y, sin duda, la paciencia suelen ser escasos, se necesitan considerar algunos atajos ante este proceso de ensayo y error. Por fortuna, el método de mínimos cuadrados ofrece un atajo. El principio o método de mínimos cuadrados elige $\hat{\beta}_1$ y $\hat{\beta}_2$ de manera que, para una muestra o conjunto de datos determinados, $\sum \hat{u}_i^2$ es la más pequeña posible. En otras palabras, para una muestra dada, proporciona valores estimados únicos de $\beta_1$ y $\beta_2$ que producen el valor más pequeño o reducido posible de $\sum \hat{u}_i^2$. ¿Cómo es esto posible? Se trata de un ejercicio sencillo

[1] Para los curiosos, estos valores se obtienen por el método de mínimos cuadrados, que explicaremos en breve. Véanse las ecuaciones (3.1.6) y (3.1.7).

de cálculo diferencial. Como se observa en el apéndice 3A, sección 3A.1, el proceso de diferenciación genera las siguientes ecuaciones para estimar $\beta_1$ y $\beta_2$:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \tag{3.1.4}$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \tag{3.1.5}$$

donde $n$ es el tamaño de la muestra. Estas ecuaciones simultáneas se conocen como **ecuaciones normales**.

Al resolver las ecuaciones normales al mismo tiempo, obtenemos

$$\begin{aligned}\hat{\beta}_2 &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ &= \frac{\sum x_i y_i}{\sum x_i^2}\end{aligned} \tag{3.1.6}$$

donde $\bar{X}$ y $\bar{Y}$ son las medias muestrales de $X$ y $Y$, y donde se definen $x_i = (X_i - \bar{X})$ y $y_i = (Y_i - \bar{Y})$. *De aquí en adelante adoptaremos la convención de utilizar letras minúsculas para representar desviaciones respecto de los valores medios.*

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \bar{Y} - \hat{\beta}_2 \bar{X}\end{aligned} \tag{3.1.7}$$

El último paso en (3.1.7) se obtiene directamente de la ecuación (3.1.4) mediante manipulación algebraica simple.

Por cierto, advierta que, con identidades algebraicas simples, la fórmula (3.1.6) para estimar $\beta_2$ se expresa también como

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2} \\ &= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2}\end{aligned} \tag{3.1.8}$$ [2]

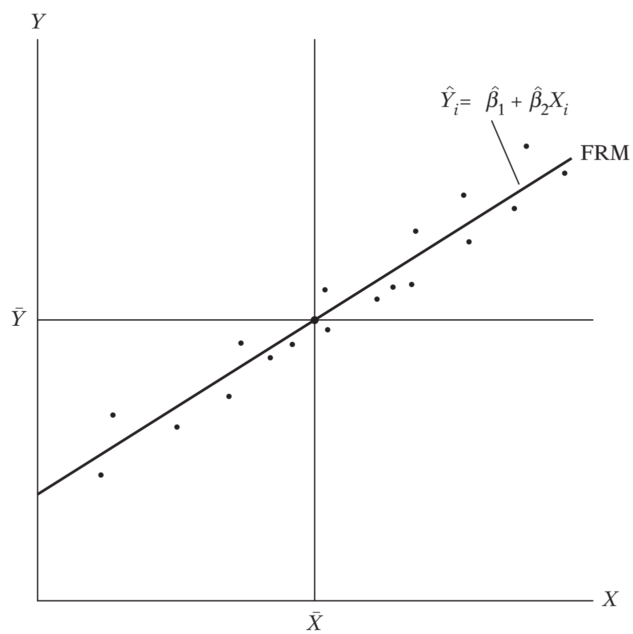
[2] *Nota* 1: $\sum x_i^2 = \sum (X_i - \bar{X})^2 = \sum X_i^2 - 2\sum X_i \bar{X} + \sum \bar{X}^2 = \sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2$, pues $\bar{X}$ es una constante. Además, aprecie que, como $\sum X_i = n\bar{X}$ y $\sum \bar{X}^2 = n\bar{X}^2$ porque $\bar{X}$ es una constante, finalmente obtenemos $\sum x_i^2 = \sum X_i^2 - n\bar{X}^2$.

*Nota* 2: $\sum x_i y_i = \sum x_i (Y_i - \bar{Y}) = \sum x_i Y_i - \bar{Y}\sum x_i = \sum x_i Y_i - \bar{Y}\sum (X_i - \bar{X}) = \sum x_i Y_i$, pues $\bar{Y}$ es una constante y la suma de las desviaciones de una variable de su valor medio [por ejemplo, $\sum (X_i - \bar{X})$ siempre es cero. De la misma manera, $\sum y_i = \sum (Y_i - \bar{Y}) = 0$.

Los estimadores obtenidos antes se conocen como **estimadores de mínimos cuadrados**, pues se derivan del principio de mínimos cuadrados. Observe las siguientes **propiedades numéricas** de los estimadores obtenidos con el método de MCO: "Propiedades numéricas son las que se mantienen como consecuencia del uso de mínimos cuadrados ordinarios, sin considerar la forma como se generaron los datos".[3] En breve consideraremos también las **propiedades estadísticas** de los estimadores MCO, es decir, propiedades "que se mantienen sólo con ciertos supuestos sobre la forma como se generaron los datos".[4] (Véase el modelo clásico de regresión lineal en la sección 3.2.)

I. Los estimadores de MCO se expresan únicamente en términos de las cantidades (es decir, $X$ y $Y$) observables (es decir, muestras). Por consiguiente, se calculan on facilidad.
II. Son **estimadores puntuales:** dada la muestra, cada estimador proporciona un solo valor (puntual) del parámetro poblacional pertinente. (En el capítulo 5 consideraremos los **estimadores por intervalos**, que proporcionan un intervalo de valores posibles para los parámetros poblacionales no conocidos.)
III. Una vez obtenidos los estimadores de MCO de los datos de la muestra, se obtiene sin problemas la línea de regresión muestral (figura 3.1). La línea de regresión así obtenida tiene las siguientes propiedades:
   1. Pasa a través de las medias muestrales de $Y$ y $X$. Esto es evidente por la ecuación (3.1.7), pues esta ecuación puede escribirse $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X}$, como se observa gráficamente en la figura 3.2.

**FIGURA 3.2**
Diagrama que muestra cómo la línea de regresión muestral pasa a través de los valores de las medias muestrales de $Y$ y $X$.




[3] Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, p. 3.

[4] *Ibid.*

2. El valor medio de $Y$ estimada $= \hat{Y}_i$ es igual al valor medio de $Y$ real para

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})\end{aligned} \tag{3.1.9}$$

Al sumar ambos lados de esta última igualdad sobre los valores muestrales y dividir por el tamaño $n$ de la muestra, obtenemos:

$$\bar{\hat{Y}} = \bar{Y} \tag{3.1.10}$$ [5]

donde se aprovecha que $\sum(X_i - \bar{X}) = 0$. (¿Por qué?)

3. El valor medio de los residuos $\hat{u}_1$ es cero. Del apéndice 3A, sección 3A.1, la primera ecuación es

$$-2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$$

Pero, como $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$, la ecuación anterior se reduce a $-2\sum \hat{u}_i = 0$, y se tiene que $\bar{\hat{u}} = 0$.[6]

Como resultado de la propiedad anterior, la regresión muestral

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{2.6.2}$$

puede definirse de una forma en la cual $Y$ y $X$ se expresan como desviaciones de sus medias. Para apreciar lo anterior, sume (2.6.2) en ambos lados para obtener

$$\begin{aligned}\sum Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \qquad \text{donde } \sum \hat{u}_i = 0\end{aligned} \tag{3.1.11}$$

Al dividir la ecuación (3.1.11) entre $n$, obtenemos

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \tag{3.1.12}$$

que es lo mismo que (3.1.7). Si restamos la ecuación (3.1.12) de la (2.6.2), obtenemos

$$Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i$$

o

$$y_i = \hat{\beta}_2 x_i + \hat{u}_i \tag{3.1.13}$$

donde $y_i$ y $x_i$, de acuerdo con lo convenido, representan desviaciones de los valores respectivos de sus medias (muestrales).

---

[5] Se debe señalar que este resultado es correcto únicamente cuando el modelo de regresión incluye el término del intercepto $\beta_1$. Como se demuestra en el **apéndice 6A, sección 6A.1**, cuando $\beta_1$ está ausente del modelo, este resultado no se da necesariamente.

[6] Este resultado también requiere que el término del intercepto $\beta_1$ esté presente en el modelo (véase el **apéndice 6A, sección 6A.1**).

La ecuación (3.1.13) se conoce como **forma de desviación**. Observe que el término del intercepto $\hat{\beta}_1$ ha desaparecido. Pero este término siempre podrá estimarse mediante (3.1.7), pues la línea de regresión muestral pasa a través de las medias muestrales de $Y$ y $X$. Una ventaja de la forma de desviación es que ésta simplifica a menudo los cálculos de las fórmulas.

Observe de paso que, en la forma de desviación, la FRM se escribe como

$$\hat{y}_i = \hat{\beta}_2 x_i \tag{3.1.14}$$

mientras que en las unidades de medición originales, dicha expresión era $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$, como en (2.6.1).

4. Los residuos $\hat{u}_i$ no están correlacionados con el valor pronosticado de $Y_i$, lo cual se verifica de la siguiente manera: con la forma de desviación se escribe

$$\begin{aligned}\sum \hat{y}_i \hat{u}_i &= \hat{\beta}_2 \sum x_i \hat{u}_i \\ &= \hat{\beta}_2 \sum x_i (y_i - \hat{\beta}_2 x_i) \\ &= \hat{\beta}_2 \sum x_i y_i - \hat{\beta}_2^2 \sum x_i^2 \\ &= \hat{\beta}_2^2 \sum x_i^2 - \hat{\beta}_2^2 \sum x_i^2 \\ &= 0\end{aligned} \tag{3.1.15}$$

donde se aprovecha que $\hat{\beta}_2 = \sum x_i y_i / \sum x_i^2$.

5. Los residuos $\hat{u}_i$ no están correlacionados con $X_i$; es decir, $\sum \hat{u}_i X_i = 0$. Esto se desprende de la ecuación (2) en el apéndice 3A, sección 3A.1.

## 3.2 Modelo clásico de regresión lineal: fundamentos del método de mínimos cuadrados

Si deseamos estimar sólo $\beta_1$ y $\beta_2$, basta el método MCO presentado de la sección anterior. Pero recuerde del capítulo 2 que en el análisis de regresión el objetivo es no sólo obtener $\hat{\beta}_1$ y $\hat{\beta}_2$, sino también inferir los verdaderos $\beta_1$ y $\beta_2$; por ejemplo, si quisiéramos saber cuán cerca están $\hat{\beta}_1$ y $\hat{\beta}_2$ de sus contrapartes en la población, o cuán cerca está $\hat{Y}_i$ de la verdadera $E(Y \mid X_i)$. Para esto no sólo se debe especificar la forma funcional del modelo, como aparece en (2.4.2), sino también hacer ciertos supuestos sobre la forma como se genera $Y_i$. Para ver por qué es necesario este requisito, observemos la FRP: $Y_i = \beta_1 + \beta_2 X_i + u_i$. Esta expresión muestra que $Y_i$ depende de $X_i$ y de $u_i$. Por consiguiente, mientras no se especifique la forma como se crean o se generan $X_i$ y *ui,* no hay manera de hacer alguna inferencia estadística sobre $Y_i$, ni tampoco, como veremos, sobre $\beta_1$ y $\beta_2$. Así, los supuestos sobre la(s) variable(s) $X_i$ y el término de error son relevantes para lograr una interpretación válida de los valores estimados de la regresión.

El **modelo de Gauss, modelo clásico** o **estándar de regresión lineal (MCRL)**, es el cimiento de la mayor parte de la teoría econométrica y plantea siete supuestos.[7] Primero los estudiaremos en el contexto del modelo de regresión con dos variables y, en el capítulo 7, se extenderán a los modelos de regresión múltiple, es decir, modelos en los cuales hay más de una regresora.

[7] Es un modelo clásico en el sentido de que Gauss lo empleó por primera vez en 1821 y desde entonces sirve como norma o patrón con el cual comparar los modelos de regresión que no satisfacen los supuestos gaussianos.

**SUPUESTO 1**

**Modelo de regresión lineal:** El modelo de regresión es **lineal en los parámetros**, aunque puede o no ser lineal en las variables. Es decir, el modelo de regresión como se muestra en la ecuación (2.4.2)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(2.4.2)}$$

Como analizaremos en el capítulo 7, este modelo puede extenderse para incluir más variables explicativas.

Ya vimos el modelo (2.4.2) en el capítulo 2. Como los modelos de regresión lineal en los parámetros son el punto de partida del MCRL, mantendremos este supuesto a lo largo del libro.[8] Recuerde que la regresada *Y* y la regresora *X* pueden no ser lineales, como vimos en el capítulo 2.

**SUPUESTO 2**

**Valores fijos de *X*, o valores de *X* independientes del término de error:** Los valores que toma la regresora *X* pueden considerarse fijos en muestras repetidas (el caso de la regresora fija), o haber sido muestreados junto con la variable dependiente *Y* (el caso de la regresora estocástica). En el segundo caso se supone que la(s) variable(s) *X* y el término de error son independientes, esto es, $\text{cov}(X_i, u_i) = 0$.

Esto puede explicarse en términos del ejemplo de la tabla 2.1 (página 35). Consideremos las diversas poblaciones de *Y* correspondientes a los niveles de ingreso en esa tabla. Al mantener el valor del ingreso *X* fijo al nivel de 80 dólares, se selecciona al azar una familia y se observa su consumo semanal *Y*, 60 dólares. Mantengamos *X* en 80 y seleccionamos aleatoriamente a otra familia, y observamos su valor *Y* de 75 dólares. En cada una de estas selecciones (es decir, muestreo repetido), el valor de *X* está fijo en 80. Se puede repetir este proceso para todos los valores de *X* de la tabla 2.1. De hecho, los datos muestrales de las tablas 2.4 y 2.5 se seleccionaron así.

¿Por qué suponemos que los valores de *X* son no estocásticos? En virtud de que en la mayoría de las ciencias sociales los datos suelen recopilarse de manera aleatoria para las variables *Y* y *X,* parece natural suponer lo contrario: que la variable *X,* lo mismo que la variable *Y,* es aleatoria o estocástica. Sin embargo, al principio suponemos que las variable(s) *X* son no estocásticas por las siguientes razones:

*Primera,* al principio, esto sirve para simplificar el análisis e introducir poco a poco al lector a las complejidades del análisis de regresión. *Segunda,* en situaciones experimentales tal vez no sea irreal suponer que los valores de *X* son fijos. Por ejemplo, un agricultor puede dividir su tierra en varias parcelas y aplicarles diferentes cantidades de fertilizante para ver el efecto en el rendimiento del cultivo. Asimismo, una tienda de departamentos puede ofrecer diferentes tasas de descuento en un producto para ver su efecto en los consumidores. En ocasiones conviene fijar los valores de *X* para un propósito específico. Supongamos que deseamos obtener el ingreso promedio semanal de los trabajadores (*Y*) con varios niveles de escolaridad (*X*), como los datos presentados en la tabla 2.6. En este caso, la variable *X* se puede considerar fija o no aleatoria. *Tercera,* como se muestra en el capítulo 13, aunque las variables *X* sean estocásticas, los resultados estadísticos de la regresión lineal basada en el caso de las regresoras fijas también

[8] Sin embargo, se presenta un análisis breve de los modelos de regresión no lineales en los parámetros, en el capítulo 14, en beneficio de los estudiantes más avanzados.

son válidos cuando las variables $X$ son aleatorias, en tanto se cumplan algunas condiciones; una de ellas es que la regresora $X$ y el término de error $u_i$ sean independientes. Como señala James Davidson: "…este modelo [es decir, el de la regresora estocástica] 'imita' al modelo de regresora fija, y […] muchas propiedades estadísticas de los mínimos cuadrados del modelo de regresora fija siguen siendo válidos".[9]

Por todas estas razones, primero analizaremos con detalle el MCRL (regresora fija). Sin embargo, en el capítulo 13 veremos el caso de las regresoras estocásticas en cierto detalle y señalaremos las ocasiones en que es necesario considerar los modelos de regresora estocástica. Por cierto, anote que si la variable $X$ es estocástica, el modelo resultante se llama **modelo neoclásico de regresión lineal (MNRL),**[10] en contraste con el MCRL, donde las $X$ se tratan como variables fijas o no aleatorias. Para efectos de análisis, denominaremos al primero **modelo de regresora estocástica,** y al segundo, **modelo de regresora fija.**

**SUPUESTO 3**

**El valor medio de la perturbación $u_i$ es igual a cero:** Dado el valor de $X_i$, la media o el valor esperado del término de perturbación aleatoria $u_i$ es cero. Simbólicamente, tenemos que

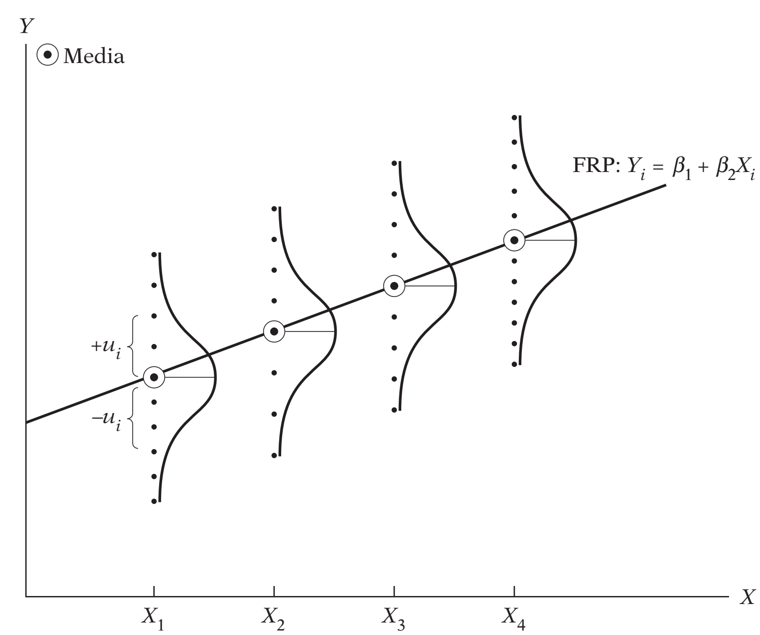
$$E(u_i \mid X_i) = 0 \qquad \textbf{(3.2.1)}$$

O, si $X$ no es estocástica,

$$E(u_i) = 0$$

El supuesto 3 establece que el valor de la media de $u_i$, que depende de las $X_i$ dadas, es cero. Geométricamente, este supuesto se representa mediante una gráfica, como en la figura 3.3, que muestra algunos valores de la variable $X$ y las poblaciones $Y$ asociadas a cada uno de ellos. Puede

**FIGURA 3.3**
Distribución condicional de las perturbaciones $u_i$.




[9] James Davidson, *Econometric Theory,* Blackwell, Reino Unido, 2000, p. 10.

[10] Término acuñado por Arthur S. Goldberger, *A Course in Econometrics,* Harvard University Press, Cambridge, Massachusetts, 1991, p. 264.

observar que cada población $Y$ correspondiente a un $X$ dado está distribuida alrededor de su media (que se representa por los puntos dentro de un círculo sobre la FRP), con algunos valores de $Y$ por encima y por debajo de ésta. Las distancias por encima y por debajo de los valores medios no son otra cosa que la $u_i$. La ecuación (3.2.1) requiere que el promedio o valor medio de estas desviaciones correspondientes a cualquier $X$ dado sea cero.

No es difícil entender este supuesto en vista de lo que vimos en la sección 2.4 (véase la ecuación 2.4.5). Lo que sostiene el supuesto 3 es que los factores no incluidos explícitamente en el modelo y, por consiguiente, incorporados en $u_i$, no afectan sistemáticamente el valor de la media de $Y$; es decir, los valores positivos de $u_i$ se cancelan con los valores negativos de $u_i$, de manera que el efecto medio o promedio sobre $Y$ es cero.[11]

Observe, por cierto, que el supuesto $E(u_i|X_i) = 0$ implica que $E(Y_i|X_i) = \beta_1 + \beta_2 X_i$. (¿Por qué?) Por consiguiente, los dos supuestos son equivalentes.

Es importante señalar que el supuesto 3 implica que no hay **sesgo de especificación** o **error de especificación** en el modelo del análisis empírico. En otras palabras, el modelo de regresión está especificado correctamente. Algunos ejemplos del error de especificación serían omitir variables explicativas importantes, incluso las variables innecesarias, o elegir una forma funcional equivocada de la relación entre las variables $Y$ y $X$. Veremos este tema a fondo en el capítulo 13.

También observe que si la media condicional de una variable aleatoria, dada otra variable aleatoria, es cero, la covarianza entre las dos variables es cero y, por tanto, las dos variables no están correlacionadas. En consecuencia, el supuesto 3 implica que $X_i$ y $u_i$ no están correlacionadas.[12]

La razón para suponer que el término de perturbación $u$ y las variables explicativas $X$ no están correlacionadas es sencilla. Cuando expresamos la FRP en la ecuación (2.4.2), supusimos que $X$ y $u$ (que representa la influencia de todas las variables omitidas) ejercen influencias independientes (y aditivas) en $Y$. Pero si $X$ y $u$ están correlacionadas, no es posible evaluar los efectos de cada una sobre $Y$. Así, si $X$ y $u$ tienen correlación positiva, $X$ aumenta cuando $u$ aumenta, y disminuye cuando $u$ disminuye. Asimismo, si $X$ y $u$ tienen correlación negativa, $X$ se incrementa cuando $u$ se reduce, y disminuye cuando $u$ aumenta. En estas situaciones es muy posible que el término de error incluya en realidad algunas variables que debieron incluirse como regresoras adicionales en el modelo. Por esta razón, el supuesto 3 es otra forma de decir que no hay error de especificación en el modelo de regresión elegido.

**SUPUESTO 4**

**Homoscedasticidad o varianza constante de $u_i$:** La varianza del término de error, o de perturbación, es la misma sin importar el valor de $X$. Simbólicamente, tenemos que

$$\begin{aligned} \text{var}(u_i) &= E[u_i - E(u_i|X_i)]^2 \\ &= E(u_i^2|X_i), \text{ por el supuesto 3} \\ &= E(u_i^2), \text{ si } X_i \text{ son variables no estocásticas} \\ &= \sigma^2 \end{aligned} \tag{3.2.2}$$

donde var significa varianza.

[11] Para una explicación más técnica de la necesidad del supuesto 3, véase E. Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1966, p. 75. Véase también el ejercicio 3.3.

[12] Sin embargo, lo contrario no es válido porque la correlación es una medida sólo de asociación lineal. Es decir, aunque $X_i$ y $u_i$ no estén correlacionadas, la media condicional de $u_i$ dada $X_i$ puede no ser cero. No obstante, si $X_i$ y $u_i$ están correlacionadas, $E(u_i|X_i)$ debe ser un número distinto de cero, lo que viola el supuesto 3. Este punto se debe a Stock y Watson. Véase James H. Stock y Mark W. Watson, *Introduction to Econometrics,* Addison-Wesley, Boston, 2003, pp. 104-105.

La ecuación (3.2.2) establece que la varianza de $u_i$ para cada $X_i$ (es decir, la varianza condicional de $u_i$) es algún número positivo constante igual a $\sigma^2$. Técnicamente, la ecuación (3.2.2) representa el supuesto de **homoscedasticidad**, o *igual* (homo) *dispersión* (cedasticidad), o *igual varianza.* La palabra proviene del verbo griego *skedanime*, que significa dispersar o esparcir. Planteado de otra forma, (3.2.2) significa que las poblaciones $Y$ correspondientes a diversos valores de $X$ tienen la misma varianza. En términos llanos, la variación alrededor de la línea de regresión (la línea de la relación promedio entre $X$ y $Y$) es la misma para todos los valores de $X$; no aumenta ni disminuye conforme varía $X$. En el diagrama de la figura 3.4 se aprecia esta situación.

En contraste, consideremos la figura 3.5, donde la varianza condicional de la población $Y$ varía con $X$. Esta situación se conoce apropiadamente como **heteroscedasticidad**, o *dispersión desigual*, o *varianza desigual*. Simbólicamente, en esta situación, la ecuación (3.2.2) se escribe como

$$\text{var}\,(u_i|X_i) = \sigma_i^2 \qquad \textbf{(3.2.3)}$$

Observe el subíndice sobre $\sigma^2$ en la ecuación (3.2.3), el cual indica que la varianza de la población $Y$ ya no es constante.

**FIGURA 3.4**
Homoscedasticidad.

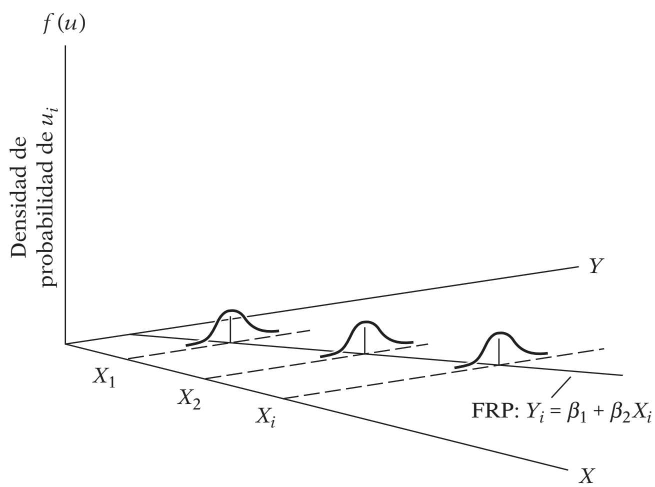
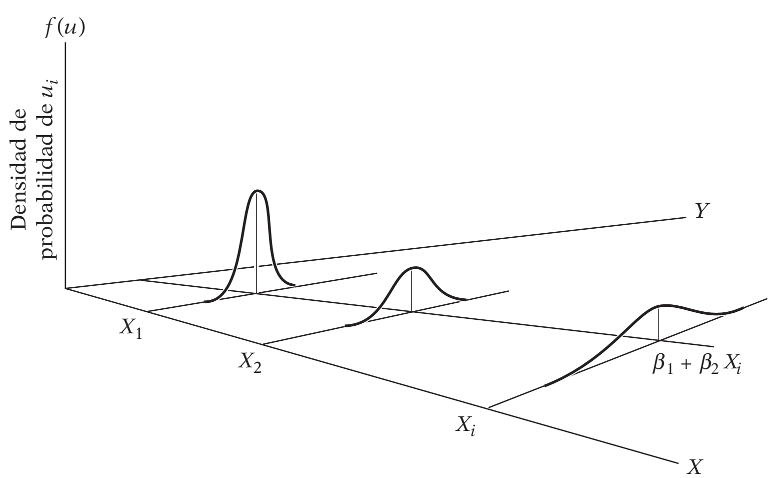



**FIGURA 3.5**
Heteroscedasticidad.



Para diferenciar claramente las dos situaciones, sea $Y$ el gasto de consumo semanal y $X$ el ingreso semanal. Las figuras 3.4 y 3.5 muestran que, a medida que el ingreso aumenta, el consumo promedio también aumenta. Pero en la figura 3.4 la varianza del consumo permanece igual para todos los niveles de ingreso, mientras que en la figura 3.5 aumenta con incrementos en el ingreso. En otras palabras, en promedio, las familias más ricas consumen más que las familias más pobres, pero hay también mayor variabilidad en el consumo que en las primeras.

Para entender el fundamento de este supuesto, observe la figura 3.5. Como lo muestra esta figura, $\text{var}(u|X_1) < \text{var}(u|X_2), \ldots, < \text{var}(u|X_i)$. Por consiguiente, lo más probable es que las observaciones de $Y$ que provienen de la población con $X = X_1$ estarían más cercanas a la FRP que las que vienen de poblaciones correspondientes a $X = X_2$, $X = X_3$, y así sucesivamente. En resumen, no todos los valores de $Y$ que corresponden a las diversas $X$ serán igualmente confiables, si juzgamos la confiabilidad por la cercanía o el alejamiento con que se distribuyan los valores de $Y$ alrededor de sus medias, es decir, los puntos sobre la FRP. Si, de hecho, éste es el caso, ¿no sería preferirible obtener muestras de las poblaciones $Y$ más cercanas a su media que de las muy dispersas? Sin embargo, actuar así restringiría la variación que se obtiene mediante los valores de $X$.

Al invocar el supuesto 4, se sostiene que en esta etapa todos los valores de $Y$ correspondientes a diversos valores de $X$ revisten la misma importancia. En el capítulo 11 veremos lo que sucede cuando se presenta heteroscedasticidad.

Note que el supuesto 4 implica que las varianzas condicionales de $Y_i$ también son homoscedásticas. Es decir,

$$\text{var}\,(Y_i|X_i) = \sigma^2 \quad \textbf{(3.2.4)}$$

Por supuesto, la *varianza incondicional* de $Y$ es $\sigma_Y^2$. Más adelante veremos la importancia de distinguir entre varianza condicional e incondicional de $Y$ (en el **apéndice A** hay detalles de varianzas condicionales e incondicionales).

**SUPUESTO 5** **No hay autocorrelación entre las perturbaciones:** Dados dos valores cualesquiera de $X$, $X_i$ y $X_j$ ($i \neq j$), la correlación entre dos $u_i$ y $u_j$ cualesquiera ($i \neq j$) es cero. En pocas palabras, estas observaciones se muestrean de manera independiente. Simbólicamente,

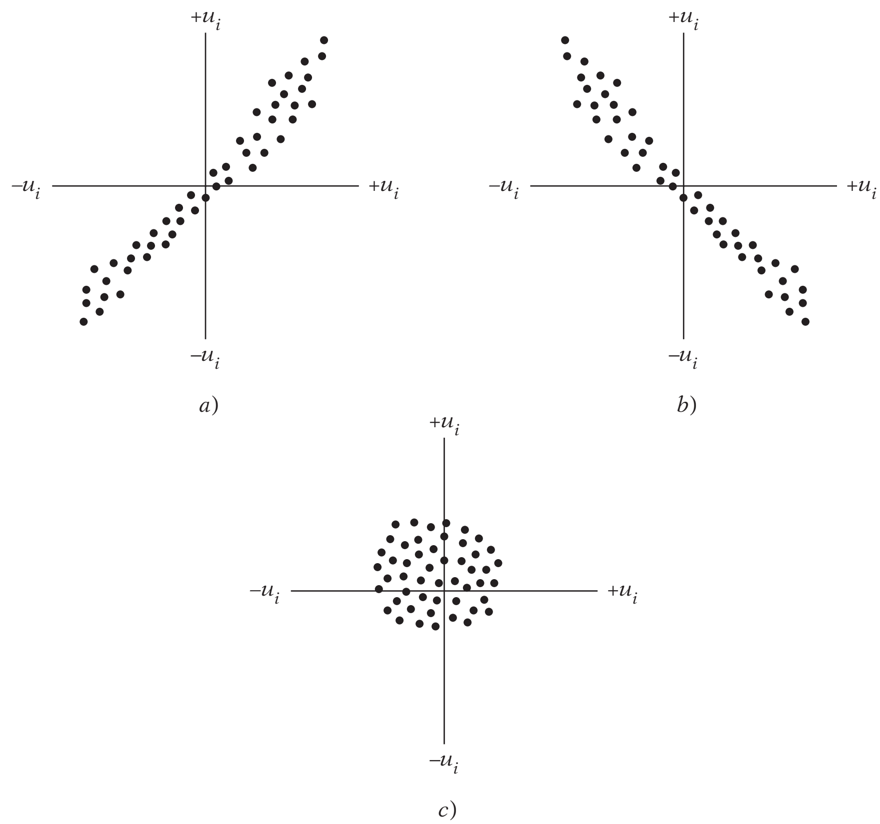
$$\text{cov}(u_i, u_j|X_i, X_j) = 0 \quad \textbf{(3.2.5)}$$
$$\text{cov}(u_i, u_j) = 0, \text{ si } X \text{ no es estocástica}$$

donde $i$ y $j$ son dos observaciones diferentes y cov significa covarianza.

En palabras, (3.2.5) postula que las perturbaciones $u_i$ y $u_j$ no están correlacionadas. Técnicamente, éste es el supuesto de **no correlación serial**, o **no autocorrelación**. Esto significa que, dado $X_i$, las desviaciones de dos valores cualesquiera de $Y$ de sus valores promedio no muestran patrones como los de la figura 3.6*a*) y *b*). En la figura 3.6*a*) se ve que las $u$ están correlacionadas positivamente, pues a una $u$ positiva sigue una $u$ positiva, o a una $u$ negativa sigue una $u$ negativa. En la figura 3.6*b*), las $u$ están **correlacionadas negativamente**, pues a una $u$ positiva sigue una $u$ negativa y viceversa.

Si las perturbaciones (desviaciones) siguen patrones sistemáticos, como los que de las figuras 3.6*a*) y *b*), hay correlación serial o autocorrelación, y lo que requiere el supuesto 5 es que dichas correlaciones estén ausentes. La figura 3.6*c*) muestra que no hay un patrón sistemático para las $u$, lo que indica cero correlación.

**FIGURA 3.6**
Patrones de correlación entre las perturbaciones: *a*) correlación serial positiva; *b*) correlación serial negativa; *c*) correlación cero.




Explicaremos con amplitud la importancia de este supuesto en el capítulo 12. No obstante, mediante la intuición, podemos analizar este supuesto de la siguiente forma. Supongamos que en la FRP ($Y_t = \beta_1 + \beta_2 X_t + u_t$) $u_t$ y $u_{t-1}$ están correlacionadas positivamente. Entonces $Y_t$ depende no sólo de $X_t$ sino también de $u_{t-1}$, pues $u_{t-1}$ determina en cierta medida a $u_t$. En esta etapa del desarrollo de la materia, al invocar el supuesto 5, se afirma que se considerará el efecto sistemático, si existe, de $X_t$ sobre $Y_t$, sin preocuparse por las demás influencias que podrían actuar sobre $Y$ como resultado de las posibles correlaciones entre las $u$. Pero, como se anota en el capítulo 12, veremos cómo incorporar en el análisis las correlaciones entre las perturbaciones, y sus consecuencias.

No obstante, debe añadirse aquí que la justificación de este supuesto depende del tipo de datos para el análisis. Si los datos son transversales y se obtienen como muestra aleatoria de la población pertinente, a menudo es posible justificar este supuesto. Sin embargo, si los datos corresponden a una serie de tiempo, es difícil mantener el supuesto de independencia, porque las observaciones sucesivas de una serie de tiempo, como el PIB, están muy correlacionadas. Analizaremos esta situación cuando estudiemos la econometría de series de tiempo, más adelante en este texto.

**SUPUESTO 6** **El número de observaciones *n* debe ser mayor que el número de parámetros por estimar:** Sucesivamente, el número de observaciones *n* debe ser mayor que el número de variables explicativas.

Este supuesto no es tan ingenuo como parece. En el ejemplo hipotético de la tabla 3.1, imaginemos que sólo había el primer par de observaciones sobre $Y$ y $X$ (4 y 1). De esta sola observación no hay forma de estimar los dos parámetros desconocidos, $\beta_1$ y $\beta_2$. Se necesitan por lo menos dos pares de observaciones para estimar dichos parámetros. En un capítulo posterior veremos la importancia crítica de este supuesto.

**SUPUESTO 7**

**La naturaleza de las variables $X$:** No todos los valores $X$ en una muestra determinada deben ser iguales. Técnicamente, var($X$) debe ser un número positivo. Además, no puede haber **valores atípicos** de la variable $X$, es decir, valores muy grandes en relación con el resto de las observaciones.

El supuesto de variabilidad en los valores de $X$ tampoco es tan ingenuo como parece. Veamos la ecuación (3.1.6). Si todos los valores de $X$ son idénticos, $X_i = \bar{X}$ (¿por qué?) y el denominador de esa ecuación será cero, lo que imposibilita la estimación de $\beta_2$ y, por consiguiente, de $\beta_1$. Por intuición, pronto advertimos la razón por la que este supuesto es importante. Observe el ejemplo del consumo familiar del capítulo 2. Si hay muy poca variación en el ingreso familiar, no será posible explicar buena parte de la variación en el consumo. El lector debe recordar que la variación tanto en $Y$ como en $X$ es esencial para utilizar el análisis de regresión como herramienta de investigación. En pocas palabras, ¡las variables deben variar!

El requisito de que no existan valores atípicos de $X$ es para evitar que los resultados de la regresión estén dominados por tales valores atípicos. Si hay algunos valores de $X$ que, por ejemplo, sean 20 veces el promedio de los valores de $X$, las líneas de regresión estimadas con o sin dichas observaciones serían muy diferentes. Con mucha frecuencia, estos valores atípicos son el resultado de errores humanos de aritmética o de mezclar muestras de diferentes poblaciones. En el capítulo 13 estudiaremos a fondo este tema.

El análisis de los supuestos en los que se basa el modelo clásico de regresión lineal ha finalizado. Es importante señalar que todos estos supuestos sólo se refieren a la FRP y no a la FRM. Sin embargo, es interesante observar que el método de mínimos cuadrados que tratamos antes tiene algunas propiedades semejantes a los supuestos que acabamos de plantear sobre la FRP. Por ejemplo, la conclusión de que $\sum \hat{u}_i = 0$ y, por tanto, $\bar{\hat{u}} = 0$, es semejante al supuesto de que $E(u_i|X_i) = 0$. Asimismo, la conclusión de que $\sum \hat{u}_i X_i = 0$ es similar al supuesto de que $\text{cov}(u_i, X_i) = 0$. Es reconfortante observar que el método de mínimos cuadrados trata de "duplicar" algunos de los supuestos impuestos a la FRP.

Desde luego, la FRM no duplica todos los supuestos del MCRL. Como mostraremos más adelante, aunque $\text{cov}(u_i, u_j) = 0 (i \neq j)$ por el supuesto, *no* es válido que la covarianza muestral $\text{cov}(\hat{u}_i, \hat{u}_j) = 0 (i \neq j)$. De hecho, habremos de demostrar que los residuos no sólo están autocorrelacionados, sino que también son heteroscedásticos (véase el capítulo 12).

## Advertencia sobre estos supuestos

La pregunta del millón de dólares es: ¿son realistas todos estos supuestos? La "realidad de los supuestos" se cuestiona desde hace muchos años en la filosofía de las ciencias. Algunos argumentan que no importa si los supuestos son realistas, sino las predicciones basadas en esos supuestos. Entre quienes apoyan la "tesis de la irrelevancia de los supuestos" sobresale Milton Friedman. Para él, la irrealidad de los supuestos es una ventaja positiva: "para que una hipótesis sea importante... debe ser descriptivamente falsa en sus supuestos".[13]

Es posible coincidir o no completamente con este punto de vista, pero recuerde que en cualquier estudio científico se plantean ciertos supuestos porque facilitan el desarrollo de la materia en pasos graduales, no porque sean necesariamente realistas en el sentido de que reproduzcan la

[13] Milton Friedman, *Essays in Positive Economics*, University of Chicago Press, Chicago, 1953, p. 14.

realidad exactamente. Como señala un autor, "... si la simplicidad es un criterio deseable de una buena teoría, todas las buenas teorías idealizan y simplifican de manera exagerada".[14]

El plan es estudiar primero a fondo las propiedades del MCRL y, luego, en capítulos posteriores, examinar a profundidad lo que sucede si no se cumple uno o más de los supuestos del MCRL. Al final de este capítulo, en la tabla 3.4, se ofrece una guía para enterarse de lo que sucede al MCRL si no se satisface un supuesto particular.

Como señalaba un colega, al revisar investigaciones ajenas, es necesario considerar si los supuestos del investigador son apropiados para los datos y para el problema. Con mucha frecuencia, la investigación publicada se basa en supuestos implícitos sobre el problema y en datos que tal vez no son correctos y producen estimaciones basadas en esos supuestos. Desde luego, el lector conocedor, consciente de estos problemas, debe adoptar una actitud escéptica hacia la investigación. Los supuestos enumerados en la tabla 3.4, por consiguiente, constituyen una lista de verificación para guiar la investigación y evaluar las investigaciones ajenas.

Con esta salvedad, ahora está listo para estudiar el MCRL. En particular, se desea encontrar las **propiedades estadísticas** de MCO comparadas con las **propiedades numéricas** puras expuestas antes. Las propiedades estadísticas de MCO se basan en los supuestos del MCRL ya estudiado, y están avaladas por el famoso **teorema de Gauss-Markov**. Pero antes de este teorema, que proporciona justificación teórica para la popularidad de MCO, se necesita considerar primero la **precisión** o los **errores estándar** de los estimados por mínimos cuadrados.

## 3.3 Precisión o errores estándar de las estimaciones de mínimos cuadrados

De las ecuaciones (3.1.6) y (3.1.7) es evidente que las estimaciones de mínimos cuadrados son función de los datos muestrales. Pero, como es probable que los datos cambien entre una muestra y otra, los valores estimados cambiarán *ipso facto*. Por consiguiente, se requiere alguna medida de "confiabilidad" o **precisión** de los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$. En estadística, la precisión de un valor estimado se mide por su error estándar (ee).[15] Dados los supuestos gaussianos, en el apéndice 3A, sección 3A.3, se muestra que los errores estándar de las estimaciones de MCO pueden obtenerse de la siguiente manera:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\operatorname{ee}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}} \tag{3.3.2}$$

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \tag{3.3.3}$$

$$\operatorname{ee}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\sigma \tag{3.3.4}$$

[14] Mark Blaug, *The Methodology of Economics: Or How Economists Explain*, 2a. ed., Cambridge University Press, Nueva York, 1992, p. 92.

[15] El **error estándar** no es otra cosa que la desviación estándar de la distribución muestral del estimador, y la distribución muestral de un estimador es tan sólo una probabilidad o distribución de frecuencias del estimador, es decir, una distribución del conjunto de valores del estimador obtenidos de todas las muestras posibles de igual tamaño de una población dada. Con las distribuciones muestrales se infieren los valores de los parámetros de la población, con base en los valores de los estimadores calculados a partir de una o más muestras (véanse detalles en el **apéndice A**).

donde var = varianza, ee = error estándar y $\sigma^2$ es la constante o varianza homoscedástica de $u_i$ del supuesto 4.

Todas las cantidades que entran en las anteriores ecuaciones, excepto $\sigma^2$, pueden estimarse a partir de los datos. Como se muestra en el apéndice 3A, sección 3A.5, la misma $\sigma^2$ se estima mediante la fórmula:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

donde $\hat{\sigma}^2$ es el estimador de MCO de la verdadera pero desconocida $\sigma^2$ , y donde la expresión $n-2$ es conocida como el **número de grados de libertad (gl)**, con $\sum \hat{u}_i^2$ como la suma de los valores residuales al cuadrado o la **suma de cuadrados de los residuos (SCR)**.[16]

Una vez conocida $\sum \hat{u}_i^2$, $\hat{\sigma}^2$ se calcula con facilidad. $\sum \hat{u}_i^2$ se obtiene de (3.1.2) o de la siguiente expresión (véase la demostración en la sección 3.5):

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

En comparación con la ecuación (3.1.2), es fácil emplear la ecuación (3.3.6), pues no requiere calcular $\hat{u}_i$ por cada observación, a pesar de la utilidad esencial de tal cálculo (como veremos en los capítulos 11 y 12).

Como

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

otra expresión para calcular $\sum \hat{u}_i^2$ es

$$\sum \hat{u}_i^2 = \sum y_i^2 - \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2} \tag{3.3.7}$$

Por cierto, note que la raíz cuadrada positiva de $\hat{\sigma}^2$

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \tag{3.3.8}$$

se conoce como el **error estándar de estimación** o el **error estándar de la regresión (ee)**. No es más que la desviación estándar de los valores $Y$ alrededor de la línea de regresión estimada, la cual suele servir como medida para resumir la "bondad del ajuste" de dicha línea, tema que analizaremos en la sección 3.5.

Antes mencionamos que, dado $X_i$, $\sigma^2$ representa la varianza (condicional) de $u_i$ y $Y_i$. Por tanto, el error estándar de la estimación también se denomina desviación estándar (condicional) de $u_i$ y $Y_i$. Sin duda, como es común, $\sigma_Y^2$ y $\sigma_Y$ representan la varianza incondicional y la desviación estándar incondicional de $Y$, respectivamente.

[16] El término **número de grados de libertad** significa el número total de observaciones en la muestra (= $n$) menos el número de restricciones (lineales) independientes o de restricciones que se les impusieron. En otras palabras, es la cantidad de observaciones independientes de un total de $n$ observaciones. Por ejemplo, para calcular la SCR (3.1.2), es necesario obtener antes $\hat{\beta}_1$ y $\hat{\beta}_2$. Por consiguiente, estas dos estimaciones imponen dos restricciones a la SCR. Son, entonces, $n-2$ las observaciones independientes, y no $n$, para calcular la SCR. Según esta lógica, en la regresión con tres variables SCR tendrá $n-3$ gl, y para el modelo de $k$ variables tendrá $n-k$ gl. **La regla general es la siguiente**: gl = ($n$ − número de parámetros estimados).

Observe las siguientes características de las varianzas (y por consiguiente, de los errores estándar) de $\hat{\beta}_1$ y $\hat{\beta}_2$.

1. La varianza de $\hat{\beta}_2$ es directamente proporcional a $\sigma^2$ pero inversamente proporcional a $\sum x_i^2$. Es decir, dada $\sigma^2$, entre más grande sea la variación en los valores $X$, menor será la varianza de $\hat{\beta}_2$ y, por tanto, mayor será la precisión con la cual estimar $\beta_2$. En resumen, dada $\sigma^2$ si hay una variación sustancial en los valores de $X$, $\beta_2$ se mide en forma más precisa que cuando las $X_i$ no varían sustancialmente. También, dado $\sum x_i^2$, entre mayor sea la varianza de $\sigma^2$, mayor será la de $\beta_2$. Advierta que a medida que aumenta el tamaño $n$ de la muestra, lo hace también el número de términos en la suma, $\sum x_i^2$. A medida que aumenta $n$, también es mayor la precisión para estimar $\beta_2$. (¿Por qué?)

2. La varianza de $\hat{\beta}_1$ es directamente proporcional a $\sigma^2$ y a $\sum X_i^2$, pero inversamente proporcional a $\sum x_i^2$ y al tamaño $n$ de la muestra.

3. Como $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores, no sólo variarán de una muestra a otra, sino también, en una muestra dada, es probable que dependan entre sí; esta dependencia se mide por la covarianza entre ellos. En el apéndice 3A, sección 3A.4, se muestra que:

$$\begin{aligned} \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) &= -\bar{X} \operatorname{var}(\hat{\beta}_2) \\ &= -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) \end{aligned} \tag{3.3.9}$$

Como var($\hat{\beta}_2$) es siempre positiva, al igual que la varianza de cualquier variable, la naturaleza de la covarianza entre $\hat{\beta}_1$ y $\hat{\beta}_2$ depende del signo de $\bar{X}$. Si $\bar{X}$ es positiva, entonces, como indica la fórmula, la covarianza será negativa. Así, si el coeficiente de la pendiente $\beta_2$ está *sobrestimado* (es decir, la pendiente es muy pronunciada), el coeficiente del intercepto $\beta_1$ estará *subestimado* (es decir, el intercepto será muy pequeño). Más adelante (sobre todo en el capítulo 10, sobre multicolinealidad), veremos la utilidad de estudiar las covarianzas entre los coeficientes estimados de regresión.

¿Cómo permiten las varianzas y los errores estándar de los coeficientes estimados de regresión evaluar la confiabilidad de estos valores estimados? Éste es un problema de la inferencia estadística, y lo trataremos en los capítulos 4 y 5.

## 3.4 Propiedades de los estimadores de mínimos cuadrados: teorema de Gauss-Markov[17]

Como ya mencionamos, dados los supuestos del modelo clásico de regresión lineal, las estimaciones de mínimos cuadrados poseen algunas propiedades ideales u óptimas, las cuales están contenidas en el famoso **teorema de Gauss-Markov**. Para entender este teorema necesitamos considerar la **propiedad del mejor estimador lineal insesgado**.[18] Como se explica en el apéndice A, se dice que un estimador, por ejemplo, el estimador de MCO $\hat{\beta}_2$, es el mejor estimador lineal insesgado (MELI) de $\beta_2$ si se cumple lo siguiente:

1. Es **lineal**, es decir, función lineal de una variable aleatoria, como la variable dependiente $Y$ en el modelo de regresión.

[17] Aunque se conoce *teorema de Gauss-Markov*, el método de Gauss de mínimos cuadrados antecede (1821) al de Markov de varianza mínima (1900).

[18] Consulte el **apéndice A,** donde se explica la importancia de los estimadores lineales y se presenta un análisis general sobre las propiedades deseables de los estimadores estadísticos.

2. Es **insesgado**, es decir, su valor promedio o esperado, $E(\hat{\beta}_2)$, es igual al valor verdadero, $\beta_2$.
3. Tiene varianza mínima dentro de la clase de todos los estimadores lineales insesgados; un estimador insesgado con varianza mínima se conoce como **estimador eficiente**.

En el contexto de regresión puede probarse que los estimadores de MCO son MELI. Ésta es la clave del famoso teorema de Gauss-Markov, el cual se puede enunciar de la siguiente forma:

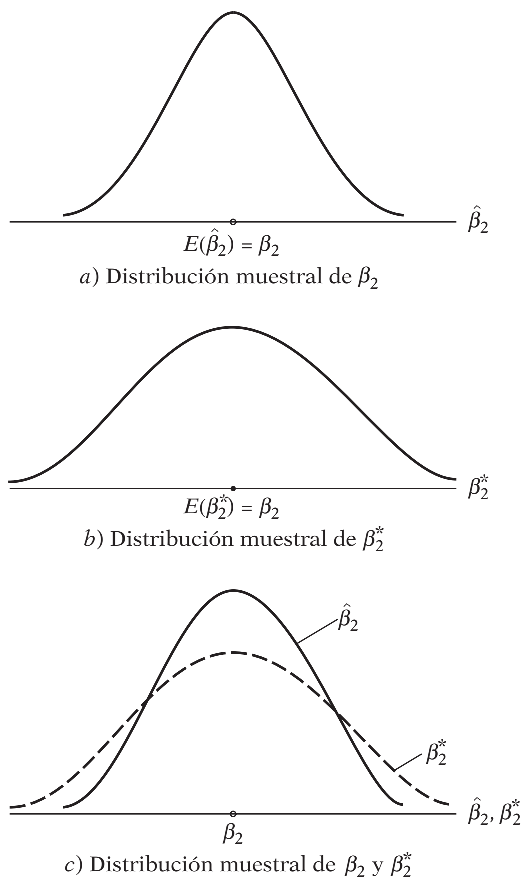
| | |
|---|---|
| ***Teorema de Gauss-Markov*** | Dados los supuestos del modelo clásico de regresión lineal, los estimadores de mínimos cuadrados, dentro de la clase de estimadores lineales insesgados, tienen varianza mínima, es decir, son MELI. |

La prueba de este teorema se presenta en el **apéndice 3A, sección 3A.6**. Conforme avancemos, percibirá con mayor claridad la trascendencia del teorema de Gauss-Markov. Es suficiente anotar aquí que el teorema tiene importancia teórica y práctica a la vez.[19]

Lo que todo esto significa se comprende con ayuda de la figura 3.7.

En la figura 3.7a se presenta la **distribución muestral** del estimador de MCO $\hat{\beta}_2$, es decir, la distribución de los valores asumidos por $\hat{\beta}_2$ en experimentos repetidos de muestreo (véase la

**FIGURA 3.7**
Distribución muestral del estimador de MCO $\hat{\beta}_2$ y el estimador alterno $\beta_2^*$.




[19] Por ejemplo, puede demostrarse que cualquier combinación lineal de las $\beta$, $(\beta_1 - 2\beta_2)$, puede estimarse por $(\hat{\beta}_1 - 2\hat{\beta}_2)$, y este estimador es MELI. Para más detalles, véase Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, pp. 401-402. Una observación sobre un asunto técnico del teorema de Gauss-Markov: sólo establece la condición suficiente (pero no la necesaria) para que los MCO sean eficientes. El autor está en deuda con Michael McAleer, de la Universidad de Western Australia, por hacerle notar este punto.

tabla 3.1). Por conveniencia, supusimos que $\hat{\beta}_2$ está distribuido simétricamente (en el capítulo 4 abundaremos al respecto). Como indica la figura, la media de los valores $\hat{\beta}_2$, $E(\hat{\beta}_2)$, es igual al verdadero $\beta_2$. En esta situación se dice que $\hat{\beta}_2$ es un *estimador insesgado* de $\beta_2$. En la figura 3.7*b*) se aprecia la distribución muestral de $\beta_2^*$, estimador alterno de $\beta_2$ obtenido con otro método (es decir, diferente al de MCO). Por conveniencia, suponga que $\beta_2^*$, al igual que $\hat{\beta}_2$, es insesgado, es decir, que su valor promedio o esperado es igual a $\beta_2$. Supongamos además que $\hat{\beta}_2$ y $\beta_2^*$ son estimadores lineales, es decir, funciones lineales de $Y$. ¿Cuál estimador escogería, $\hat{\beta}_2$ o $\beta_2^*$?

Para responder, sobreponga las dos figuras, como en la figura 3.7*c*). Es obvio que si bien $\hat{\beta}_2$ y $\beta_2^*$ son insesgados, la distribución de $\beta_2^*$ está más difusa o dispersa alrededor del valor de la media que la distribución de $\hat{\beta}_2$. En otras palabras, la varianza de $\beta_2^*$ es mayor que la varianza de $\hat{\beta}_2$. Ahora, dados dos estimadores a la vez lineales e insesgados, sería preferible el estimador con la menor varianza, porque es probable que esté más cercano a $\beta_2$, que el estimador alterno. En resumen, se escogería el estimador MELI.

El teorema de Gauss-Markov es notable, pues no hace ninguna suposición respecto de la distribución de probabilidad de la variable aleatoria $u_i$, y por consiguiente, tampoco respecto de $Y_i$ (en el siguiente capítulo abordaremos esta cuestión). En la medida en que se satisfagan los supuestos del MCRL, el teorema será válido. Como resultado, no se necesita buscar otro estimador insesgado lineal, pues no habrá otro estimador cuya varianza sea más pequeña que la del estimador de MCO. Por supuesto, si no se cumple una o más de tales suposiciones, el teorema ya no es válido. Por ejemplo, si consideramos los modelos de regresión no lineales en los parámetros (que analizaremos en el capítulo 14), quizá se obtendrían estimadores que funcionen mejor que los estimadores de MCO. Asimismo, como veremos en el capítulo sobre heteroscedasticidad, si no se cumple el supuesto sobre la varianza homoscedástica, los estimadores de MCO (aunque sean insesgados y consistentes) ya no son los estimadores de varianza mínima, incluso dentro de la clase de los estimadores lineales.

Las propiedades estadísticas que acabamos de exponer se conocen como **propiedades de muestras finitas**: estas propiedades se mantienen sin importar el tamaño de la muestra en que se basen los estimadores. Más adelante tendremos ocasión de considerar las **propiedades asintóticas**, es decir, propiedades válidas sólo si el tamaño de la muestra es muy grande (técnicamente hablando, infinito). En el **apéndice A** se presenta un análisis general de las propiedades de los estimadores con muestras finitas y muestras grandes.

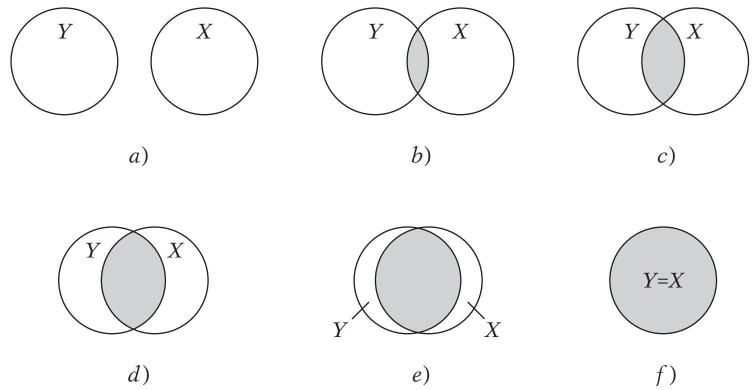
## 3.5 Coeficiente de determinación $r^2$: una medida de la "bondad del ajuste"

Hasta el momento, nuestro análisis se centró en el problema de estimar los coeficientes de regresión, sus errores estándar y algunas de sus propiedades. Veremos ahora la **bondad del ajuste** de la línea de regresión a un conjunto de datos; es decir, veremos cuán "bien" se ajusta la línea de regresión a los datos. De la figura 3.1, es claro que si todas las observaciones cayesen en la línea de regresión, obtendríamos un ajuste "perfecto", pero rara vez se presenta este caso. Por lo general hay algunas $\hat{u}_i$ positivas y algunas $\hat{u}_i$ negativas. Se tiene la esperanza de que estos residuos alrededor de la línea de regresión sean lo más pequeños posibles. El **coeficiente de determinación** $r^2$ (caso de dos variables) o $R^2$ (regresión múltiple) es una medida comprendida que dice cuán bien se ajusta la línea de regresión muestral a los datos.

Antes de mostrar cómo calcular $r^2$, consideremos una explicación heurística de $r^2$ en términos de una herramienta gráfica, conocida como el **diagrama de Venn** o de **Ballentine**, que aparece en la figura 3.8.[20]

[20] Véase Peter Kennedy, "Ballentine: A Graphical Aid for Econometrics", *Australian Economics Papers*, vol. 20, 1981, pp. 414-416. El nombre Ballentine se deriva del emblema de la conocida cerveza Ballantine con sus círculos.

**FIGURA 3.8**
Visión de Ballentine de $r^2$: *a*) $r^2 = 0$; *f*) $r^2 = 1$.




En esta figura, el círculo $Y$ representa la variación en la variable dependiente $Y$, y el círculo $X$, la variación en la variable explicativa $X$.[21] La intersección de los dos círculos (el área sombreada) indica la medida en la cual la variación en $Y$ se explica por la variación en $X$ (por ejemplo, mediante una regresión de MCO). Entre mayor sea la medida de la intersección, mayor será la variación en $Y$ que se explica por $X$. $r^2$ es tan sólo una medida numérica de esta intersección. En la figura, a medida que se va de izquierda a derecha, el área de la intersección aumenta, es decir, sucesivamente hay una proporción cada vez mayor de la variación en $Y$ que se explica por $X$. En resumen, $r^2$ aumenta. Cuando no hay intersección, obviamente $r^2$ es cero, pero cuando la intersección es completa, $r^2$ es 1, pues ciento por ciento de la variación en $Y$ se explica por $X$. Como mostraremos en breve, $r^2$ se encuentra entre 0 y 1.

Para calcular $r^2$ se procede de la siguiente forma: recuerde que

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{2.6.3}$$

o, expresado en forma de desviación,

$$y_i = \hat{y}_i + \hat{u}_i \tag{3.5.1}$$

donde se emplean (3.1.13) y (3.1.14). Al elevar al cuadrado (3.5.1) en ambos lados y sumar sobre la muestra, obtenemos

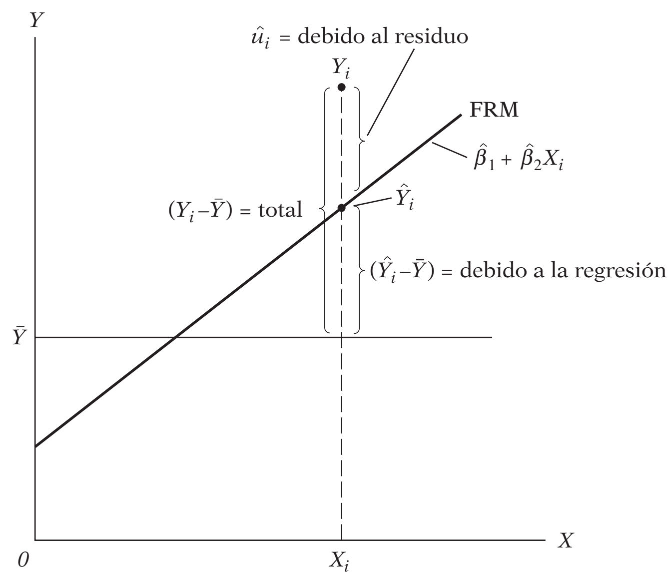
$$\begin{aligned}\sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \\ &= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2\end{aligned} \tag{3.5.2}$$

pues $\sum \hat{y}_i \hat{u}_i = 0$ (¿por qué?) y $\hat{y}_i = \hat{\beta}_2 x_i$.

Las diversas sumas de cuadrados en (3.5.2) se describen de la siguiente manera: $\sum y_i^2 = \sum (Y_i - \bar{Y})^2$ = variación total de los valores reales de $Y$ respecto de su media muestral, que puede denominarse la **suma de cuadrados total (SCT)**. $\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{\hat{Y}})^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum x_i^2$ = variación de los valores de $Y$ estimados alrededor de su media ($\bar{\hat{Y}} = \bar{Y}$), que apropiadamente puede llamarse la suma de cuadrados debida a la regresión [es decir, debida a la(s) variable(s) explicativa(s)], o explicada por ésta, o simplemente la **suma de cuadrados explicada**

[21] Los términos *variación* y *varianza* son diferentes. Variación significa la suma de los cuadrados de las desviaciones de una variable respecto del valor de su media. Varianza es la suma de los cuadrados dividida por los grados de libertad apropiados. En resumen, varianza = variación/gl.

**FIGURA 3.9**
Partición de la variación de $Y_i$ en dos componentes.




**(SCE)**. $\sum \hat{u}_i^2$ = la variación residual o **no explicada** de los valores de $Y$ alrededor de la línea de regresión, o sólo la **suma de cuadrados de los residuos (SCR)**. Así, (3.5.2) es

$$\text{SCT} = \text{SCE} + \text{SCR} \tag{3.5.3}$$

y muestra que la variación total en los valores $Y$ observados alrededor del valor de su media puede dividirse en dos partes, una atribuible a la línea de regresión y la otra a fuerzas aleatorias, pues no todas las observaciones $Y$ caen sobre la línea ajustada. Geométricamente, tenemos la figura 3.9.

Ahora, al dividir la ecuación (3.5.3) entre la SCT en ambos lados, se obtiene

$$\begin{aligned} 1 &= \frac{\text{SCE}}{\text{SCT}} + \frac{\text{SCR}}{\text{SCT}} \\ &= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} \end{aligned} \tag{3.5.4}$$

Ahora, definimos $r^2$ como

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{SCE}}{\text{SCT}} \tag{3.5.5}$$

o también como

$$\begin{aligned} r^2 &= 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} \\ &= 1 - \frac{\text{SCR}}{\text{SCT}} \end{aligned} \tag{3.5.5a}$$

La cantidad $r^2$ así definida se conoce como **coeficiente de determinación** (muestral), y es la medida más común de la bondad del ajuste de una línea de regresión. Verbalmente, $r^2$ *mide la proporción o el porcentaje de la variación total en Y explicada por el modelo de regresión.*

Pueden observarse dos propiedades de $r^2$:

1. Es una cantidad no negativa. (¿Por qué?)
2. Sus límites son $0 \le r^2 \le 1$. Un $r^2$ de 1 significa un ajuste perfecto, es decir, $\hat{Y}_i = Y_i$ por cada $i$. Por otra parte, un $r^2$ de cero significa que no hay relación alguna entre la variable regresada y la variable regresora (es decir, $\hat{\beta}_2 = 0$). En este caso, como indica (3.1.9), $\hat{Y}_i = \hat{\beta}_1 = \bar{Y}$, es decir, la mejor predicción de cualquier valor de $Y$ es simplemente el valor de su media. En esta situación, por consiguiente, la línea de regresión será horizontal al eje $X$.

A pesar de que $r^2$ puede calcularse directamente a partir de su definición en (3.5.5), su valor se obtiene más rápido con la siguiente fórmula:

$$\begin{aligned} r^2 &= \frac{\text{SCE}}{\text{SCT}} \\ &= \frac{\sum \hat{y}_i^2}{\sum y_i^2} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} \\ &= \hat{\beta}_2^2 \left( \frac{\sum x_i^2}{\sum y_i^2} \right) \end{aligned} \tag{3.5.6}$$

Si dividimos el numerador y el denominador de (3.5.6) por el tamaño $n$ de la muestra (o $n-1$, si la muestra es pequeña), obtenemos:

$$r^2 = \hat{\beta}_2^2 \left( \frac{S_x^2}{S_y^2} \right) \tag{3.5.7}$$

donde $S_y^2$ y $S_x^2$ son las varianzas muestrales de $Y$ y $X$, respectivamente.

Como $\hat{\beta}_2 = \sum x_i y_i \big/ \sum x_i^2$, la ecuación (3.5.6) también se expresa como

$$r^2 = \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2 \sum y_i^2} \tag{3.5.8}$$

expresión fácil de calcular.

Con la definición de $r^2$, SCE y SCR, explicadas antes, se expresan de la siguiente forma:

$$\begin{aligned} \text{SCE} &= r^2 \cdot \text{SCT} \\ &= r^2 \sum y_i^2 \end{aligned} \tag{3.5.9}$$

$$\begin{aligned} \text{SCR} &= \text{SCT} - \text{SCE} \\ &= \text{SCT}(1 - \text{SCE}/\text{SCT}) \\ &= \sum y_i^2 \cdot (1 - r^2) \end{aligned} \tag{3.5.10}$$

Por consiguiente, escribimos

$$\begin{gathered} \text{SCT} = \text{SCE} + \text{SCR} \\ \sum y_i^2 = r^2 \sum y_i^2 + (1 - r^2) \sum y_i^2 \end{gathered} \tag{3.5.11}$$

expresión que nos será muy útil más adelante.

Una cantidad estrechamente relacionada con $r^2$ pero conceptualmente muy diferente es el **coeficiente de correlación**, el cual, como mencionamos en el capítulo 1, es una medida del grado de asociación entre dos variables. Se calcula a partir de

$$r = \pm\sqrt{r^2} \tag{3.5.12}$$

o de su definición

$$\begin{aligned} r &= \frac{\sum x_i y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}} \\ &= \frac{n\sum X_iY_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n\sum Y_i^2 - \left(\sum Y_i\right)^2\right]}} \end{aligned} \tag{3.5.13}$$

que se conoce como **coeficiente de correlación muestral**.[22]

Algunas propiedades de $r$ son las siguientes (véase la figura 3.10):

1. Puede tener signo positivo o negativo, según el signo del término en el numerador de (3.5.13), el cual mide la *covariación* muestral de dos variables.

2. Cae entre los límites de $-1$ y $+1$; es decir, $-1 \leq r \leq 1$.

3. Es simétrico por naturaleza; es decir, el coeficiente de correlación entre $X$ y $Y$ ($r_{XY}$) es el mismo que entre $Y$ y $X$ ($r_{YX}$).

4. Es independiente del origen y de la escala; es decir, si definimos $X_i^* = aX_i + C$ y $Y_i^* = bY_i + d$, donde $a > 0$, $b > 0$, y $c$ y $d$ son constantes, entonces $r$ entre $X^*$ y $Y^*$ es igual a $r$ entre las variables originales $X$ y $Y$.

5. Si $X$ y $Y$ son estadísticamente independientes (véase la definición en el **apéndice A**), el coeficiente de correlación entre ellas es cero; pero si $r = 0$, esto no significa que las dos variables sean independientes. En otras palabras, **una correlación igual a cero no necesariamente implica independencia**. [Véase la figura 3.10*h*).]

6. Es una medida de *asociación lineal* o *dependencia lineal* solamente; su uso en la descripción de relaciones no lineales no tiene significado. Así, en la figura 3.10*h*), $Y = X^2$ es una relación exacta y a pesar de ello $r$ es cero. (¿Por qué?)

7. Aunque es una medida de asociación lineal entre dos variables, esto no implica necesariamente alguna relación causa-efecto, como mencionamos en el capítulo 1.

En el contexto de la regresión, $r^2$ es una medida con más significado que $r$, pues la primera indica la proporción de la variación en la variable dependiente explicada por la(s) variable(s) explicativa(s) y, por consiguiente, constituye una medida global del grado en que la variación en una variable determina la variación en la otra. La segunda no tiene tal valor.[23] Además, como veremos, la interpretación de $r (= R)$ en un modelo de regresión múltiple es de valor dudoso. Sin embargo, tendremos más que decir sobre $r^2$ en el capítulo 7.

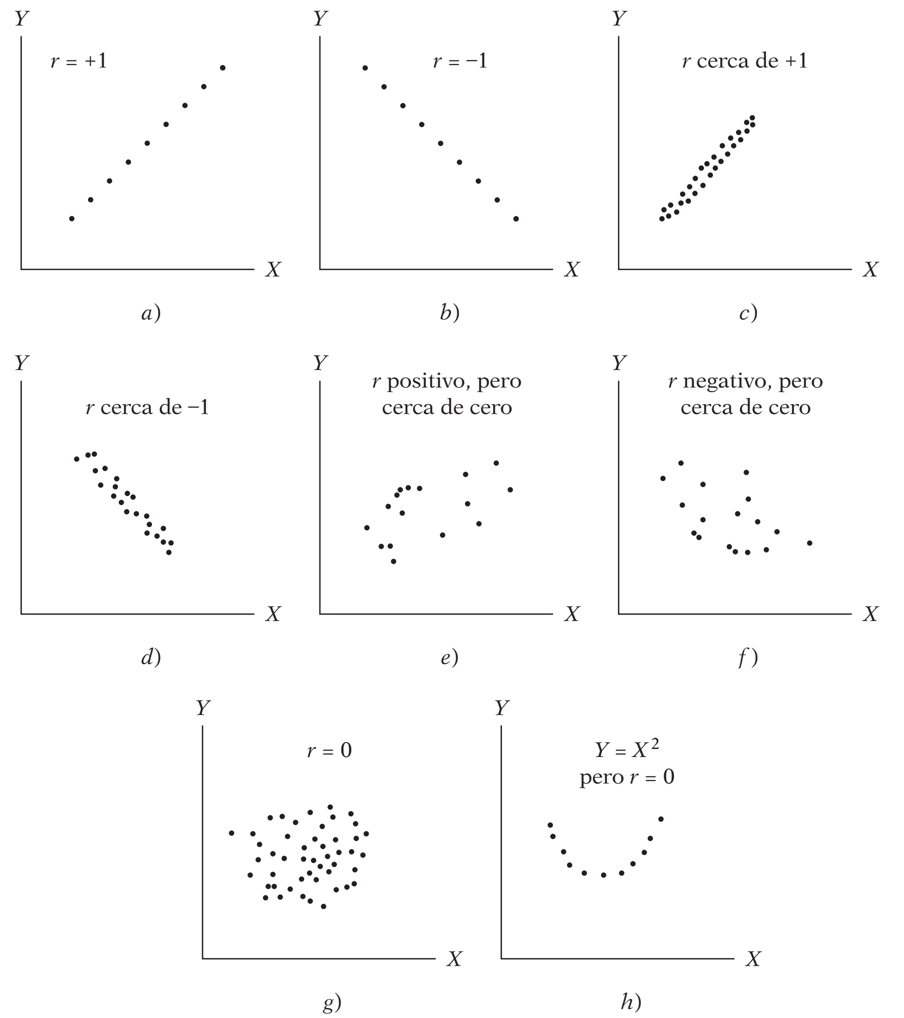
Observe que la $r^2$ definida antes *también puede calcularse como el coeficiente de correlación entre la $Y_i$ real y la $Y_i$ estimada, a saber, $\hat{Y}_i$, elevado al cuadrado.* Es decir, con (3.5.13), se escribe

$$r^2 = \frac{\left[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})\right]^2}{\sum(Y_i - \bar{Y})^2 \sum(\hat{Y}_i - \bar{Y})^2}$$

[22] El coeficiente de correlación poblacional, denotado por $\rho$, se define en el **apéndice A**.

[23] En el proceso de creación de modelos de regresión, la teoría indicará la dirección de causalidad entre $Y$ y $X$, la cual, en el contexto de los modelos uniecuacionales, suele presentarse de $X$ a $Y$.

**FIGURA 3.10**
Patrones de correlación (adaptado de Henri Theil, *Introduction to Econometrics,* Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, p. 86).




Es decir,

$$r^2 = \frac{\left(\sum y_i \hat{y}_i\right)^2}{\left(\sum y_i^2\right)\left(\sum \hat{y}_i^2\right)} \tag{3.5.14}$$

donde $Y_i = Y$ real, $\hat{Y}_i = Y$ estimada y $\bar{Y} = \bar{\hat{Y}} =$ media de $Y$. Para la prueba, consulte el ejercicio 3.15. La expresión (3.5.14) justifica la descripción de $r^2$ como medida de la bondad del ajuste, pues señala qué tan cerca están los valores de $Y$ estimados de sus valores observados.

## 3.6 Ejemplo numérico

Para ilustrar la teoría econométrica hasta el momento, consideraremos los datos de la tabla 2.6, que relacionan el salario promedio por hora ($Y$) y los años de escolaridad ($X$). La economía laboral básica indica que, entre muchas variables, la escolaridad es un determinante importante de los salarios.

En la tabla 3.2 se proporcionan los datos primarios que se necesitan para estimar el efecto cuantitativo de la escolaridad en los salarios.

**TABLA 3.2**
**Datos primarios basados en la tabla 2.6**

| Obs | $Y$ | $X$ | $x$ | $y$ | $x_i^2$ | $y_i x_i$ |
|---|---|---|---|---|---|---|
| 1 | 4.4567 | 6 | −6 | **−4.218** | 36 | 25.308 |
| 2 | 5.77 | 7 | −5 | **−2.9047** | 25 | 14.5235 |
| 3 | 5.9787 | 8 | −4 | **−2.696** | 16 | 10.784 |
| 4 | 7.3317 | 9 | −3 | **−1.343** | 9 | 4.029 |
| 5 | 7.3182 | 10 | −2 | **−1.3565** | 4 | 2.713 |
| 6 | 6.5844 | 11 | −1 | **−2.0903** | 1 | 2.0903 |
| 7 | 7.8182 | 12 | 0 | **−0.8565** | 0 | 0 |
| 8 | 7.8351 | 13 | 1 | **−0.8396** | 1 | −0.8396 |
| 9 | 11.0223 | 14 | 2 | **2.3476** | 4 | 4.6952 |
| 10 | 10.6738 | 15 | 3 | **1.9991** | 9 | 5.9973 |
| 11 | 10.8361 | 16 | 4 | **2.1614** | 16 | 8.6456 |
| 12 | 13.615 | 17 | 5 | **4.9403** | 25 | 24.7015 |
| 13 | 13.531 | 18 | 6 | **4.8563** | 36 | 29.1378 |
| Suma | 112.7712 | 156 | 0 | **0** | 182 | 131.7856 |

| Obs | $X_i^2$ | $Y_i^2$ | $\hat{Y}_i$ | $\hat{u}_i = Y_i - \hat{Y}$ | $\hat{u}_i^2$ |
|---|---|---|---|---|---|
| 1 | 36 | 19.86217 | 4.165294 | 0.291406 | 0.084917 |
| 2 | 49 | 33.2929 | 4.916863 | 0.853137 | 0.727843 |
| 3 | 64 | 35.74485 | 5.668432 | 0.310268 | 0.096266 |
| 4 | 81 | 53.75382 | 6.420001 | 0.911699 | 0.831195 |
| 5 | 100 | 53.55605 | 7.17157 | 0.14663 | 0.0215 |
| 6 | 121 | 43.35432 | 7.923139 | −1.33874 | 1.792222 |
| 7 | 144 | 61.12425 | 8.674708 | −0.85651 | 0.733606 |
| 8 | 169 | 61.38879 | 9.426277 | −1.59118 | 2.531844 |
| 9 | 196 | 121.4911 | 10.17785 | 0.844454 | 0.713103 |
| 10 | 225 | 113.93 | 10.92941 | −0.25562 | 0.065339 |
| 11 | 256 | 117.4211 | 11.68098 | −0.84488 | 0.713829 |
| 12 | 289 | 185.3682 | 12.43255 | 1.182447 | 1.398181 |
| 13 | 324 | 183.088 | 13.18412 | 0.346878 | 0.120324 |
| Suma | 2 054 | 1 083.376 | 112.7712 | ≈0 | 9.83017 |

*Nota:*

$$x_i = X_i - \bar{X};\ y_i = Y_i = \bar{Y}$$

$$\hat{\beta}_2 = \frac{\Sigma y_i x_i}{\Sigma x_i^2} = \frac{131.7856}{182.0} = 0.7240967$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 8.674708 - 0.7240967x12 = -0.01445$$

$$\hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{n-2} = \frac{9.83017}{11} = 0.893652;\ \hat{\sigma} = 0.945332$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\Sigma x_i^2} = \frac{0.893652}{182.0} = 0.004910;\ \text{ee}(\hat{\beta}_2) = \sqrt{0.00490} = 0.070072$$

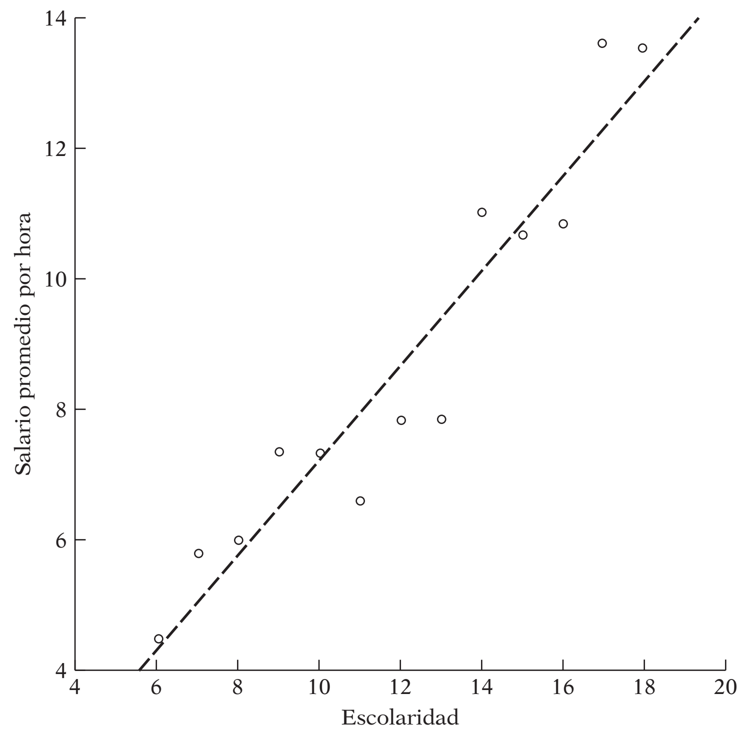
$$r^2 = 1 - \frac{\Sigma \hat{u}_i^2}{\Sigma (Y_i - \bar{Y})^2} = 1 - \frac{9.83017}{105.1188} = 0.9065$$

$$r = \sqrt{r^2} = 0.9521$$

$$\text{var}(\hat{\beta}_1) = \frac{\Sigma x_i^2}{n \Sigma x_i^2} = \frac{2\,054}{13(182)} = 0.868132;$$

$$\text{ee}(\hat{\beta}_1) = \sqrt{0.868132} = 0.9317359$$

**FIGURA 3.11**
Línea de regresión estimada para los datos de salarios y escolaridad de la tabla 2.6.




Con los datos presentados en esta tabla, la línea de regresión estimada se obtiene como sigue:

$$\hat{Y}_i = -0.0144 + 0.7240X_i \tag{3.6.1}$$

Geométricamente, la línea de regresión estimada se muestra en la figura 3.11.

Como sabemos, cada punto de la línea de regresión da una estimación del valor medio de $Y$ que corresponde al valor seleccionado de $X$; es decir, $\hat{Y}_i$ es una estimación de $E(Y \mid X_i)$. El valor de $\hat{\beta}_2 = 0.7240$, que mide la pendiente de la línea, indica que, dentro del intervalo muestral de $X$ entre 6 y 18 años de escolaridad, a medida que el valor de $X$ aumenta 1 dólar, el incremento estimado en el salario promedio por hora es cercano a 72 centavos de dólar. Es decir, cada año adicional de escolaridad, en promedio, produce aumentos en los salarios por hora de alrededor de 72 centavos de dólar.

El valor de $\hat{\beta}_1 = -0.0144$, que es el intercepto de la línea, indica el nivel promedio de los salarios cuando el nivel de escolaridad es cero. Esta interpretación literal del intercepto en el presente caso no tiene sentido. ¿Cómo podría haber salarios negativos? Como veremos a lo largo de este libro, ocurre con mucha frecuencia que el término del intercepto no tiene un significado viable ni práctico. Además, en la muestra no se contempla el nivel cero de escolaridad. Como veremos en el capítulo 5, el valor observado del intercepto no es estadísticamente diferente de cero.

El valor de $r^2$ cercano a 0.90 indica que el nivel de escolaridad explica alrededor de 90% de la variación del salario por hora. Si consideramos que $r^2$ puede tener un valor máximo de 1, la línea de regresión se ajusta muy bien a los datos. El coeficiente de correlación, $r = 0.9521$, muestra que los salarios y la escolaridad tienen una alta correlación positiva.

Antes de dejar este ejemplo, observe que el modelo es muy sencillo. La teoría económica laboral indica que, aparte de la escolaridad, las variables como género, raza, ubicación, sindicatos e idioma son también factores importantes en la determinación de los salarios por hora. Después de estudiar la regresión múltiple en los capítulos 7 y 8, consideraremos un modelo más amplio de determinación de los salarios.

## 3.7 Ejemplos ilustrativos

**EJEMPLO 3.1**
*Relación consumo-ingreso en Estados Unidos, 1960-2005*

Retomemos los datos sobre ingreso y consumo de la tabla I.1, en la Introducción. Ya presentamos estos datos en la figura I.3, junto con la línea de regresión estimada en la ecuación (I.3.3). Ahora proporcionamos los resultados subyacentes de la regresión de MCO que se obtuvieron con *Eviews* 6. Observe que $Y$ = gasto de consumo personal (GCP) y $X$ = producto interno bruto (PIB), ambos en miles de millones de dólares de 2000. En este ejemplo, los datos son de series de tiempo.

$$\hat{Y}_t = -299.5913 + 0.7218X_t \tag{3.7.1}$$

$$\text{var}(\hat{\beta}_1) = 827.4195 \qquad \text{ee}(\hat{\beta}_1) = 28.7649$$
$$\text{var}(\hat{\beta}_2) = 0.0000195 \qquad \text{ee}(\hat{\beta}_2) = 0.004423$$
$$r^2 = 0.9983 \qquad \hat{\sigma}^2 = 73.56689$$

La ecuación (3.7.1) es la función de consumo agregada keynesiana (es decir, para la economía en su conjunto). Como muestra esta ecuación, la **propensión marginal a consumir (PMC)** es de cerca de 0.72, lo que indica que si el ingreso real se incrementa un dólar, el gasto promedio de consumo personal aumenta casi 72 centavos. Según la teoría keynesiana, se espera que la PMC se sitúe entre 0 y 1.

El valor del intercepto en este ejemplo es negativo y no tiene ninguna interpretación económica viable. De manera textual, significa que si el valor del PIB fuera cero, el nivel promedio del consumo personal sería un valor negativo de alrededor de 299 000 millones de dólares.

El valor de $r^2$ de 0.9983 significa que más o menos 99% de la variación en el consumo personal se explica por la variación en el PIB. Este valor es muy alto, si se considera que $r^2$ puede valer cuando mucho 1. Como veremos a lo largo de esta obra, en las regresiones basadas en datos de series de tiempo por lo general se obtienen valores altos de $r^2$. Explicaremos las razones de este fenómeno en el capítulo que trata sobre la autocorrelación, y también en el capítulo sobre econometría de series de tiempo.

**EJEMPLO 3.2**
*Gasto alimentario en India*

Consulte los datos de la tabla 2.8 del ejercicio 2.15. Los datos se refieren a una muestra de 55 familias rurales de India. La variable dependiente (regresada) en este ejemplo es el gasto en alimentos y la independiente (regresora) es el gasto total, una aproximación del ingreso (ambas cifras se dan en rupias). Los datos de este ejemplo son, por tanto, *transversales*.

Con base en los datos proporcionados, obtenemos la siguiente regresión:

$$\widehat{\text{GasAl}}_i = 94.2087 + 0.4368\ \text{GasTot}_i \tag{3.7.2}$$

$$\text{var}(\hat{\beta}_1) = 2\,560.9401 \qquad \text{ee}(\hat{\beta}_1) = 50.8563$$
$$\text{var}(\hat{\beta}_2) = 0.0061 \qquad \text{ee}(\hat{\beta}_2) = 0.0783$$
$$r^2 = 0.3698 \qquad \hat{\sigma}^2 = 4\,469.6913$$

En la ecuación (3.7.2) se observa que si el gasto total se incrementa una rupia, en promedio, el gasto en alimentos aumenta casi 44 paisas (1 rupia = 100 paisas). Si el gasto total fuera nulo (cero), el gasto promedio en alimentos sería de más o menos 94 rupias. De nuevo, tal interpretación mecánica del intercepto no tendría ningún sentido. Sin embargo, en este ejemplo se puede argumentar que aunque el gasto total fuera nulo (por ejemplo, debido a la pérdida del trabajo), la gente podría mantener un nivel mínimo de gasto en comida si pide dinero prestado o recurre a sus ahorros.

El valor de $r^2$ de casi 0.37 significa que sólo 37% de la variación en el gasto alimentario se explica por el gasto total. Esto puede parecer un valor más bien bajo, pero, como veremos después, en los datos transversales suelen obtenerse valores bajos de $r^2$, quizá debido a la diversidad de unidades de la muestra. Analizaremos este tema en el capítulo sobre heteroscedasticidad (véase el capítulo 11).

**EJEMPLO 3.3**
*Demanda de teléfonos celulares y computadoras personales en relación con el ingreso per cápita*

La tabla 3.3 presenta datos sobre el número de suscriptores de teléfonos celulares y el número de computadoras personales (PC), ambos por cada 100 personas, y el ingreso *per cápita* ajustado por el poder adquisitivo en dólares para una muestra de 34 países. Por tanto, se trata de datos transversales. Estos datos corresponden a 2003 y se obtuvieron del *Statistical Abstract of the United States,* 2006.

Aunque los teléfonos celulares y las computadoras personales son muy comunes en Estados Unidos, no ocurre lo mismo en muchos países. Para ver si el ingreso *per cápita* es un factor que influye en el uso de teléfonos celulares y PC, se regresó cada uno de estos medios de comunicación sobre el ingreso *per cápita* con la muestra de 34 países. Los resultados son los siguientes:

**TABLA 3.3**
**Número de suscriptores de teléfonos celulares por cada cien personas y número de computadoras personales por cada cien personas e ingreso *per cápita* en países seleccionados, correspondientes a 2003**

Fuente: *Statistical Abstract of the United States,* 2006, tabla 1364 para los datos sobre teléfonos celulares y computadoras, y tabla 1327 para ingreso *per cápita* ajustado por poder adquisitivo.

| País | Teléfono celular | PC | Ingreso *per cápita* (EU$) |
|---|---|---|---|
| Alemania | 78.52 | 48.47 | 27 610 |
| Arabia Saudita | 32.11 | 13.67 | 13 230 |
| Argentina | 17.76 | 8.2 | 11 410 |
| Australia | 71.95 | 60.18 | 28 780 |
| Bélgica | 79.28 | 31.81 | 28 920 |
| Brasil | 26.36 | 7.48 | 7 510 |
| Bulgaria | 46.64 | 5.19 | 75.4 |
| Canadá | 41.9 | 48.7 | 30 040 |
| China | 21.48 | 2.76 | 4 980 |
| Colombia | 14.13 | 4.93 | 6 410 |
| Ecuador | 18.92 | 3.24 | 3 940 |
| Egipto | 8.45 | 2.91 | 3 940 |
| España | 91.61 | 19.6 | 22 150 |
| Estados Unidos | 54.58 | 40.57 | 37 750 |
| Francia | 69.59 | 34.71 | 27 640 |
| Grecia | 90.23 | 8.17 | 19 900 |
| Guatemala | 13.15 | 1.44 | 4 090 |
| Hungría | 76.88 | 10.84 | 13 840 |
| India | 2.47 | 0.72 | 2 880 |
| Indonesia | 8.74 | 1.19 | 3 210 |
| Italia | 101.76 | 23.07 | 26 830 |
| Japón | 67.9 | 38.22 | 28 450 |
| México | 29.47 | 8.3 | 8 980 |
| Países Bajos | 76.76 | 46.66 | 28 560 |
| Pakistán | 1.75 | 0.42 | 2 040 |
| Polonia | 45.09 | 14.2 | 11 210 |
| Reino Unido | 91.17 | 40.57 | 27 690 |
| República Checa | 96.46 | 17.74 | 15 600 |
| Rusia | 24.93 | 8.87 | 8 950 |
| Sudáfrica | 36.36 | 7.26 | 10 130 |
| Suecia | 98.05 | 62.13 | 26 710 |
| Suiza | 84.34 | 70.87 | 32 220 |
| Tailandia | 39.42 | 3.98 | 7 450 |
| Venezuela | 27.3 | 6.09 | 4 750 |

*Nota:* Los datos sobre teléfonos celulares y computadoras personales son por cada 100 personas.

**Demanda de teléfonos celulares.** Sea $Y$ = número de suscriptores de teléfonos celulares y $X$ = ingreso *per cápita* ajustado por poder adquisitivo. Con esta base se obtuvo la siguiente regresión.

$$\begin{aligned} \hat{Y}_i &= 14.4773 + 0.0022X_i \\ \text{ee}(\hat{\beta}_1) &= 6.1523; \quad \text{ee}(\hat{\beta}_2) = 0.00032 \\ r^2 &= 0.6023 \end{aligned} \tag{3.7.3}$$

El coeficiente de pendiente indica que si el ingreso *per cápita* aumenta, por ejemplo, 1 000 dólares en promedio, el número de suscriptores de teléfonos celulares aumentará alrededor de 2.2 por cada 100 personas. El valor del intercepto de 14.47 indica que, aunque el ingreso *per cápita* sea cero, el número promedio de suscriptores de teléfonos celulares es de alrededor de 14 por cada 100 personas. Una vez más, es posible que esta interpretación no tenga mucho sentido, pues en la muestra no se incluye ningún país con ingreso *per cápita* cero. El valor de $r^2$ es moderadamente alto. Sin embargo, observe que la muestra incluye varios países con diferentes niveles de ingreso. En una muestra tan diversa, no se esperaría un valor de $r^2$ muy alto.

Después de estudiar el capítulo 5, mostraremos que los errores estándar estimados que se obtienen con la ecuación 3.7.3 sirven para evaluar el significado estadístico de los coeficientes estimados.

**Demanda de computadoras personales.** Aunque los precios de las computadoras personales han disminuido considerablemente con el transcurso de los años, las PC aún no están presentes en todas partes. Un determinante importante de la demanda de computadoras personales es el ingreso personal. Otro determinante es el precio, pero no se dispone de datos comparativos de los precios de PC en los países de la muestra.

Si $Y$ denota el número de PC y $X$ el ingreso *per cápita*, se obtiene la siguiente demanda "parcial" de PC (parcial porque no se cuenta con datos de precios comparativos ni datos sobre otras variables que podrían afectar la demanda de PC).

$$\begin{aligned} \hat{Y}_i &= -6.5833 + 0.0018X_i \\ \text{ee}(\hat{\beta}_1) &= 2.7437; \quad \text{ee}(\hat{\beta}_2) = 0.00014 \\ r^2 &= 0.8290 \end{aligned} \tag{3.7.4}$$

Como indican estos resultados, el ingreso *per cápita* tiene relación positiva con la demanda de PC. Después de estudiar el capítulo 5 veremos que, en términos estadísticos, el ingreso *per cápita* es un determinante importante de la demanda de PC. El valor negativo del intercepto en el presente caso no tiene un significado práctico. Pese a la diversidad de la muestra, el valor estimado de $r^2$ es muy alto. La interpretación del coeficiente de pendiente es que si el ingreso *per cápita* aumenta, por ejemplo, 1 000 dólares en promedio, la demanda de computadoras personales aumentará en alrededor de 2 unidades por cada 100 personas.

Aunque el uso de las computadoras personales se extiende con rapidez, hay muchos países en los que todavía se usan computadoras centrales (*mainframes*). Por tanto, el uso total de computadoras en esos países puede ser mucho más elevado de lo que indica la venta de PC.

## 3.8 Una observación sobre los experimentos Monte Carlo

En este capítulo vimos que, conforme a los supuestos del MCRL, los estimadores de mínimos cuadrados tienen ciertas características estadísticas deseables que se resumen en la propiedad MELI. En el apéndice de este capítulo comprobamos esta propiedad más formalmente. Pero en la práctica, ¿cómo saber si se mantiene la propiedad MELI? Por ejemplo, ¿cómo se puede averiguar si los estimadores de MCO son insesgados? La respuesta proviene de los llamados experimentos **Monte Carlo**, los cuales son, en esencia, experimentos de muestreo o de simulación en computadora.

Para introducir las ideas básicas, consideremos la FRP de dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3.8.1}$$

Un experimento Monte Carlo se realiza de la siguiente forma:

1. Supongamos que los valores verdaderos de los parámetros son los siguientes: $\beta_1 = 20$ y $\beta_2 = 0.6$.
2. Escogemos el tamaño de la muestra, por ejemplo, $n = 25$.
3. Fijamos los valores de $X$ para cada observación. En total se tendrán 25 valores de $X$.
4. Supongamos que se consulta una tabla de números aleatorios, elegimos 25 valores y los denominamos $u_i$ (hoy en día, la mayoría de los paquetes estadísticos tienen generadores de números aleatorios integrados).[24]
5. Como *conocemos* $\beta_1$, $\beta_2$, $X_i$ y $u_i$, con la ecuación (3.8.1) obtenemos 25 valores de $Y_i$.
6. Ahora, con los 25 valores de $Y_i$ generados de esa forma, efectuamos la regresión de estos valores sobre los 25 valores de $X$ seleccionados en el paso 3, y así se obtienen los estimadores de mínimos cuadrados $\hat{\beta}_1$ y $\hat{\beta}_2$.
7. Supongamos que repite este experimento 99 veces, siempre con los mismos valores de $\beta_1$, $\beta_2$ y $X$. Sin duda, los valores $u_i$ variarán de un experimento a otro. Por consiguiente, en total se tienen 100 experimentos, para generar así 100 valores para cada $\beta_1$ y $\beta_2$. (En la práctica se realizan muchos experimentos de este tipo; en ocasiones llegan a 1 000 o 2 000.)
8. Tomamos los promedios de estos 100 valores estimados y los denominamos $\bar{\hat{\beta}}_1$ y $\bar{\hat{\beta}}_2$.
9. Si estos valores promedio son más o menos los mismos que los valores verdaderos de $\beta_1$ y $\beta_2$ supuestos en el paso 1, mediante este experimento Monte Carlo se "establece" que, en efecto, los estimadores de mínimos cuadrados son insesgados. Recuerde que, según el MCRL, $E(\hat{\beta}_1) = \beta_1$ y $E(\hat{\beta}_2) = \beta_2$.

Estos pasos caracterizan la naturaleza general de los experimentos Monte Carlo. Tales experimentos son comunes al estudiar las propiedades estadísticas de diversos métodos de estimación de parámetros poblacionales. Son en particular útiles para estudiar el comportamiento de los estimadores en muestras pequeñas, o finitas. Estos experimentos son también un medio excelente de demostración del concepto de **muestreo repetido**, que es la base de la mayor parte de la inferencia estadística clásica, como veremos en el capítulo 5. Se presentarán diversos ejemplos de los experimentos Monte Carlo en forma de ejercicios para realizar en clase (véase el ejercicio 3.27).

## Resumen y conclusiones

Los temas y conceptos importantes de este capítulo se resumen de la siguiente forma.

1. El marco básico del análisis de regresión es el **MCRL**.
2. El MCRL se basa en un conjunto de supuestos.
3. Con base en estos supuestos, los estimadores de mínimos cuadrados adquieren ciertas propiedades resumidas en el teorema de Gauss-Markov, el cual plantea que dentro de la clase de estimadores lineales insesgados, los estimadores de mínimos cuadrados tienen una varianza mínima. En resumen, son MELI.
4. La *precisión* de los estimadores de MCO se mide por sus **errores estándar**. En los capítulos 4 y 5 veremos que los errores estándar permiten hacer inferencias sobre los parámetros poblacionales, los coeficientes $\beta$.
5. La bondad del ajuste general del modelo de regresión se mide con el **coeficiente de determinación**, $r^2$. Éste indica qué proporción de la variación en la variable dependiente, o variable regresada, se explica por la variable explicativa, o regresora. $r^2$ se sitúa entre 0 y 1; entre más cerca esté de 1, mejor será el ajuste.

[24] En la práctica se supone que $u_i$ sigue una cierta distribución de probabilidad, digamos, normal, con algunos parámetros (es decir, la media y la varianza). Una vez especificados los parámetros, con paquetes estadísticos se generan con facilidad las $u_i$.

6. Un concepto relacionado con el coeficiente de determinación es el **coeficiente de correlación**, $r$. Es una medida de *asociación lineal* entre dos variables y su valor se encuentra entre $-1$ y $+1$.
7. El MCRL es un concepto o abstracción teórica, pues se basa en un conjunto de supuestos que pueden ser rigurosos o "poco reales". Sin embargo, tal abstracción es a menudo necesaria en las etapas iniciales del estudio de cualquier disciplina. Una vez dominado el MCRL, se puede saber lo que sucede si uno o más de sus supuestos no se satisfacen. La primera parte de este libro se dedica al estudio del MCRL. Las demás partes del libro se consideran refinaciones del MCRL. La tabla 3.4 señala el mapa del camino.

**TABLA 3.4**
**¿Qué sucede si se violan los supuestos del MCRL?**

| Número del supuesto | Tipo de violación | ¿Dónde estudiarlo? |
|---|---|---|
| 1 | No linealidad en parámetros | Capítulo 14 |
| 2 | Regresora(s) estocástica(s) | Capítulo 13 |
| 3 | Media de $u_i$ distinta de cero | Introducción a la parte II |
| 4 | Heteroscedasticidad | Capítulo 11 |
| 5 | Perturbaciones autocorrelacionadas | Capítulo 12 |
| 6 | Observaciones muestrales menores que el número de regresoras | Capítulo 10 |
| 7 | Variabilidad insuficiente en las regresoras | Capítulo 10 |
| 8 | Multicolinealidad* | Capítulo 10 |
| 9 | Sesgo de especificación* | Capítulos 13 y 14 |
| 10** | No normalidad de las perturbaciones | Capítulo 13 |

*Estos supuestos se presentan en el capítulo 7, cuando se estudie el modelo de regresión múltiple.
***Nota*: El supuesto de que las perturbaciones $u_i$ están normalmente distribuidas no forma parte del MCRL. Véase más al respecto en el capítulo 4.

# EJERCICIOS

## Preguntas

3.1. Dados los supuestos en la columna 1 de la siguiente tabla, demuestre que los supuestos en la columna 2 son equivalentes.

**Supuestos del modelo clásico**

| (1) | (2) |
|---|---|
| $E(u_i \mid X_i) = 0$ | $E(Y_i \mid X_i) = \beta_2 + \beta_2 X$ |
| $\text{cov}\,(u_i, u_j) = 0 \; i \neq j$ | $\text{cov}\,(Y_i, Y_j) = 0 \; i \neq j$ |
| $\text{var}\,(u_i \mid X_i) = \sigma^2$ | $\text{var}\,(Y_i \mid X_i) = \sigma^2$ |

3.2. Demuestre que los valores estimados $\hat{\beta}_1 = 1.572$ y $\hat{\beta}_2 = 1.357$ del primer experimento de la tabla 3.1 son en realidad los estimadores de MCO.

3.3. De acuerdo con Malinvaud (véase la nota 11), el supuesto de que $E(u_i \mid X_i) = 0$ es muy importante. Para ver esto, considere la FRP: $Y = \beta_1 + \beta_2 X_i + u_i$. Ahora considere dos situaciones: *i*) $\beta_1 = 0$, $\beta_2 = 1$ y $E(u_i) = 0$; y *ii*) $\beta_1 = 1$, $\beta_2 = 2$ y $E(u_i) = (X_i - 1)$. Ahora obtenga la esperanza de la FRP condicional sobre $X$ en los dos casos anteriores y vea si está de acuerdo con Malinvaud sobre la significancia del supuesto $E(u_i \mid X_i) = 0$.

3.4. Considere la regresión muestral

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Impuestas las restricciones *i*) $\sum \hat{u}_i = 0$ y *ii*) $\sum \hat{u}_i X_i = 0$, obtenga los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$, y demuestre que son idénticos a los estimadores de mínimos cuadrados dados en (3.1.6) y (3.1.7). Este método de obtención de estimadores se denomina **principio de analogía**. Dé una justificación intuitiva de la imposición de las restricciones *i*) y *ii*). (*Sugerencia:* recuerde los supuestos del MCRL sobre $u_i$.) Además, tenga en cuenta que el principio de analogía para la estimación de parámetros desconocidos se llama también **método de momentos**, en el cual los momentos muestrales (por ejemplo, la media muestral) sirven para estimar los momentos poblacionales (por ejemplo, la media poblacional). Como se anota en el **apéndice A**, un **momento** es un resumen estadístico de una distribución de probabilidad, como el valor esperado y la varianza.

3.5. Demuestre que $r^2$ definido en (3.5.5) varía entre 0 y 1. Utilice la desigualdad de Cauchy-Schwarz, la cual establece que para dos variables aleatorias *X* y *Y* cualesquiera se cumple la siguiente relación:

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

3.6. Sean $\hat{\beta}_{YX}$ y $\hat{\beta}_{XY}$ las pendientes en la regresión de *Y* sobre *X* y de *X* sobre *Y*, respectivamente. Demuestre que

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

donde *r* es el coeficiente de correlación entre *X* y *Y*.

3.7. Suponga en el ejercicio 3.6 que $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$. ¿Tiene importancia entonces si se hace la regresión *Y* sobre *X*, o *X* sobre *Y*? Explique detalladamente.

3.8. El coeficiente de correlación de orden de Spearman, $r_s$, se define de la siguiente forma:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

donde $d =$ la diferencia en las posiciones o lugares asignados al mismo individuo o fenómeno, y $n =$ número de individuos o de fenómenos ordenados. Obtenga $r_s$ a partir de la *r* definida en (3.5.13). *Sugerencia:* ordene los valores de *X* y *Y* de 1 hasta *n*. Note que la suma de todas las posiciones de *X* y *Y* es $n(n+1)/2$ y, por consiguiente, sus medias son $(n+1)/2$.

3.9. Considere las siguientes formulaciones de la FRP de dos variables:

$$\text{Modelo I:} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\text{Modelo II:} \quad Y_i = \alpha_1 + \alpha_2(X_i - \bar{X}) + u_i$$

*a*) Encuentre los estimadores de $\beta_1$ y $\alpha_1$. ¿Son idénticos? ¿Sus varianzas son idénticas?

*b*) Encuentre los estimadores de $\beta_2$ y $\alpha_2$. ¿Son idénticos? ¿Sus varianzas son idénticas?

*c*) ¿Cuál es la ventaja, si acaso, del modelo II sobre el modelo I?

3.10. Suponga que realiza la siguiente regresión:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$$

donde, como siempre, $y_i$ y $x_i$ son desviaciones de sus respectivos valores medios. ¿Cuál será el valor de $\hat{\beta}_1$? ¿Por qué? ¿Será $\hat{\beta}_2$ igual al obtenido de la ecuación (3.1.6)? ¿Por qué?

3.11. Sea $r_1$ = coeficiente de correlación entre $n$ pares de valores $(Y_i, X_i)$ y $r_2$ = coeficiente de correlación entre $n$ pares de valores $(aX_i + b, cY_i + d)$, donde $a, b, c$ y $d$ son constantes. Demuestre que $r_1 = r_2$ y de aquí *establezca el principio de que el coeficiente de correlación es invariante frente a cambios de escala o cambios de origen.*

*Sugerencia:* aplique la definición de $r$ dada en (3.5.13).

*Nota:* las operaciones $aX_i$, $X_i + b$, y $aX_i + b$ se conocen respectivamente por los nombres de *cambios de escala, cambios de origen* y *cambios de escala y de origen*.

3.12. Si $r$, el coeficiente de correlación entre $n$ pares de valores $(X_i, Y_i)$, es positivo, determine si las siguientes afirmaciones son ciertas o falsas:

*a*) $r$ entre $(-X_i, -Y_i)$ es también positivo.

*b*) $r$ entre $(-X_i, Y_i)$ y aquel entre $(X_i, -Y_i)$ pueden ser positivos o negativos.

*c*) Los dos coeficientes de pendiente, $\beta_{yx}$ y $\beta_{xy}$ son positivos, donde $\beta_{yx}$ = coeficiente de la pendiente en la regresión de $Y$ sobre $X$ y $\beta_{xy}$ = coeficiente de la pendiente en la regresión de $X$ sobre $Y$.

3.13. Si $X_1$, $X_2$ y $X_3$ son variables no correlacionadas y cada una tiene la misma desviación estándar, demuestre que el coeficiente de correlación entre $X_1 + X_2$ y $X_2 + X_3$ es igual a $\frac{1}{2}$. ¿Por qué el coeficiente de correlación no es cero?

3.14. En la regresión $Y_i = \beta_1 + \beta_2 X_i + u_i$, suponga que se *multiplica* cada valor de $X$ por una constante, 2, por ejemplo. ¿Cambiará esto los residuos y los valores ajustados de $Y$? Explique. ¿Qué sucede si se *agrega* un valor constante, por ejemplo, 2, a cada valor de $X$?

3.15. Demuestre que (3.5.14) mide en realidad el coeficiente de determinación. *Sugerencia:* aplique la definición de $r$ dada en (3.5.13) y recuerde que $\sum y_i \hat{y}_i = \sum(\hat{y}_i + \hat{u}_i)\hat{y}_i = \sum \hat{y}_i^2$, y recuerde (3.5.6).

3.16. *Dé razones* por las que los siguientes enunciados son verdaderos, falsos o ambiguos.

*a*) Como la correlación entre dos variables, $Y$ y $X$, puede variar de –1 a +1, esto significa que $\text{cov}(Y, X)$ también está dentro de esos límites.

*b*) Si la correlación entre dos variables es cero, esto quiere decir que no existe ninguna relación entre las dos variables.

*c*) Si se hace la regresión de $Y_i$ sobre $\hat{Y}_i$ (es decir, la $Y$ real sobre la $Y$ estimada), el valor del intercepto y de la pendiente serán respectivamente 0 y 1.

3.17. *Regresión sin regresora*. Suponga que se le proporciona el siguiente modelo: $Y_i = \beta_1 + u_i$. Con los MCO, determine el estimador de $\beta_1$. ¿Cuál es su varianza y su SCR? ¿La estimación $\beta_1$ tiene algún sentido intuitivo? Ahora considere el modelo de dos variables $Y_i = \beta_1 + \beta_2 X_i + u_i$. ¿Vale la pena añadir $X_i$ al modelo? Si no es así, ¿por qué molestarse con el análisis de regresión?

## Ejercicios empíricos

3.18. En la tabla 3.5 se dan las posiciones o lugares de 10 estudiantes en los exámenes intermedios y finales de estadística. Calcule el coeficiente de correlación de orden de Spearman e interprételo:

**TABLA 3.5**

| | Estudiante | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Posición** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** |
| Intermedios | 1 | 3 | 7 | 10 | 9 | 5 | 4 | 8 | 2 | 6 |
| Finales | 3 | 2 | 8 | 7 | 9 | 6 | 5 | 10 | 1 | 4 |

3.19. *Relación entre el tipo de cambio nominal y los precios relativos*. A partir de las observaciones anuales de 1985 a 2005, se obtuvieron los siguientes resultados de regresión, donde $Y$ = tipo de cambio del dólar canadiense respecto del dólar estadounidense (CD/\$) y $X$ = razón entre el índice de precios al consumidor estadounidense y el índice de precios al consumidor canadiense; es decir, $X$ representa los precios relativos en ambos países:

$$\hat{Y}_t = -0.912 + 2.250X_t \qquad r^2 = 0.440$$
$$\text{ee} = \qquad\qquad 0.096$$

*a*) Interprete esta regresión. ¿Cómo interpretaría $r^2$?

*b*) ¿El valor positivo de $X_t$ tiene sentido económico? ¿En qué teoría económica se basa?

*c*) Suponga que se fuera a redefinir $X$ como la razón entre el IPC canadiense respecto del IPC estadounidense. ¿Lo anterior haría cambiar el signo de $X$? ¿Por qué?

3.20. La tabla 3.6 proporciona datos sobre los índices de producción por hora ($X$) y la compensación real por hora ($Y$) de los negocios y sectores no agrícolas de la economía de Estados Unidos de 1960 a 2005. El año base para los índices es 1992 = 100; además, los índices se ajustan por estacionalidad.

*a*) Grafique por separado $Y$ respecto de $X$ para los dos sectores.

*b*) ¿En qué teoría económica se basa la relación entre ambas variables? ¿El diagrama de dispersión apoya esta teoría?

*c*) Estime la regresión MCO de $Y$ sobre $X$. Guarde los resultados para una revisión posterior, cuando estudie el capítulo 5.

3.21. De una muestra de 10 observaciones se obtuvieron los siguientes resultados:

$$\sum Y_i = 1\ 110 \qquad \sum X_i = 1\ 700 \qquad \sum X_i Y_i = 205\ 500$$
$$\sum X_i^2 = 322\ 000 \qquad \sum Y_i^2 = 132\ 100$$

con el coeficiente de correlación $r = 0.9758$. Pero al verificar estos cálculos se descubrió que se registraron dos pares de observaciones:

| Y | X | | Y | X |
|---|---|---|---|---|
| 90 | 120 | en lugar de | 80 | 110 |
| 140 | 220 | | 150 | 210 |

¿Cuál será el efecto de este error en $r$? Obtenga la $r$ correcta.

3.22. La tabla 3.7 presenta los datos sobre el precio del oro, el índice de precios al consumidor (IPC) y el índice de la Bolsa de Valores de Nueva York (BVNY) de Estados Unidos de 1974 a 2006. El índice de la BVNY incluye la mayor parte de las acciones registradas, las cuales ascienden a más de 1 500.

*a*) En el mismo diagrama de dispersión, grafique los precios del oro, el IPC y el índice de la BVNY.

*b*) Se supone que una inversión es una protección contra la inflación si su precio o la tasa de rendimiento se mantiene por lo menos al ritmo de la inflación. Para probar esta hipótesis, suponga que se decide ajustar el siguiente modelo, suponiendo que el gráfico de los puntos dispersos en *a*) indica que esto es lo apropiado:

$$\text{Precio del oro}_t = \beta_1 + \beta_2\,\text{IPC}_t + u_t$$
$$\text{Índice BVNY}_t = \beta_1 + \beta_2\,\text{IPC}_t + u_t$$

**TABLA 3.6**
**Productividad y datos relacionados, sector de negocios 1960-2005 (cifras de índices, 1992 = 100; datos trimestrales ajustados por estacionalidad)**

Fuente: *Economic Report of the President*, 2007, tabla 49.

| | Producción por hora de todas las personas[1] | | Remuneración real por hora[2,3] | |
|---|---|---|---|---|
| Año | Sector de negocios | Sector de negocios no agrícolas | Sector de negocios | Sector de negocios no agrícolas |
| 1960 | 48.9 | 51.9 | 60.8 | 63.3 |
| 1961 | 50.6 | 53.5 | 62.5 | 64.8 |
| 1962 | 52.9 | 55.9 | 64.6 | 66.7 |
| 1963 | 55.0 | 57.8 | 66.1 | 68.1 |
| 1964 | 56.8 | 59.6 | 67.7 | 69.3 |
| 1965 | 58.8 | 61.4 | 69.1 | 70.5 |
| 1966 | 61.2 | 63.6 | 71.7 | 72.6 |
| 1967 | 62.5 | 64.7 | 73.5 | 74.5 |
| 1968 | 64.7 | 66.9 | 76.2 | 77.1 |
| 1969 | 65.0 | 67.0 | 77.3 | 78.1 |
| 1970 | 66.3 | 68.0 | 78.8 | 79.2 |
| 1971 | 69.0 | 70.7 | 80.2 | 80.7 |
| 1972 | 71.2 | 73.1 | 82.6 | 83.2 |
| 1973 | 73.4 | 75.3 | 84.3 | 84.7 |
| 1974 | 72.3 | 74.2 | 83.3 | 83.8 |
| 1975 | 74.8 | 76.2 | 84.1 | 84.5 |
| 1976 | 77.1 | 78.7 | 86.4 | 86.6 |
| 1977 | 78.5 | 80.0 | 87.6 | 88.0 |
| 1978 | 79.3 | 81.0 | 89.1 | 89.6 |
| 1979 | 79.3 | 80.7 | 89.3 | 89.7 |
| 1980 | 79.2 | 80.6 | 89.1 | 89.6 |
| 1981 | 80.8 | 81.7 | 89.3 | 89.8 |
| 1982 | 80.1 | 80.8 | 90.4 | 90.8 |
| 1983 | 83.0 | 84.5 | 90.3 | 90.9 |
| 1984 | 85.2 | 86.1 | 90.7 | 91.1 |
| 1985 | 87.1 | 87.5 | 92.0 | 92.2 |
| 1986 | 89.7 | 90.2 | 94.9 | 95.2 |
| 1987 | 90.1 | 90.6 | 95.2 | 95.5 |
| 1988 | 91.5 | 92.1 | 96.5 | 96.7 |
| 1989 | 92.4 | 92.8 | 95.0 | 95.1 |
| 1990 | 94.4 | 94.5 | 96.2 | 96.1 |
| 1991 | 95.9 | 96.1 | 97.4 | 97.4 |
| 1992 | 100.0 | 100.0 | 100.0 | 100.0 |
| 1993 | 100.4 | 100.4 | 99.7 | 99.5 |
| 1994 | 101.3 | 101.5 | 99.0 | 99.1 |
| 1995 | 101.5 | 102.0 | 98.7 | 98.8 |
| 1996 | 104.5 | 104.7 | 99.4 | 99.4 |
| 1997 | 106.5 | 106.4 | 100.5 | 100.3 |
| 1998 | 109.5 | 109.4 | 105.2 | 104.9 |
| 1999 | 112.8 | 112.5 | 108.0 | 107.5 |
| 2000 | 116.1 | 115.7 | 112.0 | 111.5 |
| 2001 | 119.1 | 118.6 | 113.5 | 112.8 |
| 2002 | 124.0 | 123.5 | 115.7 | 115.1 |
| 2003 | 128.7 | 128.0 | 117.7 | 117.1 |
| 2004 | 132.7 | 131.8 | 119.0 | 118.2 |
| 2005 | 135.7 | 134.9 | 120.2 | 119.3 |

[1] Producción se refiere al producto interno bruto real en el sector.
[2] Sueldos y salarios de los empleados, más la contribución de los patrones al seguro social y los planes privados de prestaciones.
[3] Remuneración por hora dividida entre el índice de precios al consumidor para todos los consumidores urbanos en los trimestres recientes.

**TABLA 3.7**
**Precios del oro, índice de la Bolsa de Valores de Nueva York e Índice de Precios al Consumidor de Estados Unidos, 1974-2006**

| Año | Precio del oro | BVNY | IPC |
|---|---|---|---|
| 1974 | 159.2600 | 463.5400 | 49.30000 |
| 1975 | 161.0200 | 483.5500 | 53.80000 |
| 1976 | 124.8400 | 575.8500 | 56.90000 |
| 1977 | 157.7100 | 567.6600 | 60.60000 |
| 1978 | 193.2200 | 567.8100 | 65.20000 |
| 1979 | 306.6800 | 616.6800 | 72.60000 |
| 1980 | 612.5600 | 720.1500 | 82.40000 |
| 1981 | 460.0300 | 782.6200 | 90.90000 |
| 1982 | 375.6700 | 728.8400 | 96.50000 |
| 1983 | 424.3500 | 979.5200 | 99.60000 |
| 1984 | 360.4800 | 977.3300 | 103.9000 |
| 1985 | 317.2600 | 1 142.970 | 107.6000 |
| 1986 | 367.6600 | 1 438.020 | 109.6000 |
| 1987 | 446.4600 | 1 709.790 | 113.6000 |
| 1988 | 436.9400 | 1 585.140 | 118.3000 |
| 1989 | 381.4400 | 1 903.360 | 124.0000 |
| 1990 | 383.5100 | 1 939.470 | 130.7000 |
| 1991 | 362.1100 | 2 181.720 | 136.2000 |
| 1992 | 343.8200 | 2 421.510 | 140.3000 |
| 1993 | 359.7700 | 2 638.960 | 144.5000 |
| 1994 | 384.0000 | 2 687.020 | 148.2000 |
| 1995 | 384.1700 | 3 078.560 | 152.4000 |
| 1996 | 387.7700 | 3 787.200 | 156.9000 |
| 1997 | 331.0200 | 4 827.350 | 160.5000 |
| 1998 | 294.2400 | 5 818.260 | 163.0000 |
| 1999 | 278.8800 | 6 546.810 | 166.6000 |
| 2000 | 279.1100 | 6 805.890 | 172.2000 |
| 2001 | 274.0400 | 6 397.850 | 177.1000 |
| 2002 | 309.7300 | 5 578.890 | 179.9000 |
| 2003 | 363.3800 | 5 447.460 | 184.0000 |
| 2004 | 409.7200 | 6 612.620 | 188.9000 |
| 2005 | 444.7400 | 7 349.000 | 195.3000 |
| 2006 | 603.4600 | 8 357.990 | 201.6000 |

3.23. La tabla 3.8 proporciona datos sobre el producto interno bruto (PIB) de Estados Unidos de 1959 a 2005.

*a*) Grafique los datos del PIB en dólares actuales y constantes (es decir, de 2000) respecto del tiempo.

*b*) Sea $Y$ el PIB y $X$ el tiempo (medido cronológicamente, empezando con 1 para 1959, 2 para 1960, hasta 47 para 2005), observe si el siguiente modelo se ajusta a los datos del PIB:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Estime este modelo para el PIB en dólares constantes y actuales.

*c*) ¿Cómo interpretaría $\beta_2$?

*d*) Si existiera una diferencia entre el valor estimado de $\beta_2$ para el PIB en dólares actuales y el estimado para el PIB en dólares constantes, ¿qué la explicaría?

*e*) Con base en sus resultados, ¿qué puede decir sobre la naturaleza de la inflación en Estados Unidos en el periodo de muestra?

**TABLA 3.8**
**Producto interno bruto nominal y real, Estados Unidos, 1959-2005 (miles de millones de dólares, salvo si se indica otra cosa; datos trimestrales a tasas anuales ajustadas por estacionalidad; producto interno bruto real [PIBR] en miles de millones de dólares de 2000 ajustados en el tiempo por la inflación)**

| Año | PIBN | PIBR | Año | PIBN | PIBR |
|---|---|---|---|---|---|
| 1959 | 506.6 | 2 441.3 | 1983 | 3 536.7 | 5 423.8 |
| 1960 | 526.4 | 2 501.8 | 1984 | 3 933.2 | 5 813.6 |
| 1961 | 544.7 | 2 560.0 | 1985 | 4 220.3 | 6 053.7 |
| 1962 | 585.6 | 2 715.2 | 1986 | 4 462.8 | 6 263.6 |
| 1963 | 617.7 | 2 834.0 | 1987 | 4 739.5 | 6 475.1 |
| 1964 | 663.6 | 2 998.6 | 1988 | 5 103.8 | 6 742.7 |
| 1965 | 719.1 | 3 191.1 | 1989 | 5 484.4 | 6 981.4 |
| 1966 | 787.8 | 3 399.1 | 1990 | 5 803.1 | 7 112.5 |
| 1967 | 832.6 | 3 484.6 | 1991 | 5 995.9 | 7 100.5 |
| 1968 | 910.0 | 3 652.7 | 1992 | 6 337.7 | 7 336.6 |
| 1969 | 984.6 | 3 765.4 | 1993 | 6 657.4 | 7 532.7 |
| 1970 | 1 038.5 | 3 771.9 | 1994 | 7 072.2 | 7 835.5 |
| 1971 | 1 127.1 | 3 898.6 | 1995 | 7 397.7 | 8 031.7 |
| 1972 | 1 238.3 | 4 105.0 | 1996 | 7 816.9 | 8 328.9 |
| 1973 | 1 382.7 | 4 341.5 | 1997 | 8 304.3 | 8 703.5 |
| 1974 | 1 500.0 | 4 319.6 | 1998 | 8 747.0 | 9 066.9 |
| 1975 | 1 638.3 | 4 311.2 | 1999 | 9 268.4 | 9 470.3 |
| 1976 | 1 825.3 | 4 540.9 | 2000 | 9 817.0 | 9 817.0 |
| 1977 | 2 030.9 | 4 750.5 | 2001 | 10 128.0 | 9 890.7 |
| 1978 | 2 294.7 | 5 015.0 | 2002 | 10 469.6 | 10 048.8 |
| 1979 | 2 563.3 | 5 173.4 | 2003 | 10 960.8 | 10 301.0 |
| 1980 | 2 789.5 | 5 161.7 | 2004 | 11 712.5 | 10 703.5 |
| 1981 | 3 128.4 | 5 291.7 | 2005 | 12 455.8 | 11 048.6 |
| 1982 | 3 255.0 | 5 189.3 | | | |

Fuente: *Economic Report of the President*, 2007 tablas B-1 y B-2.

3.24. Con los datos de la tabla I.1 de la Introducción, verifique la ecuación (3.7.1).

3.25. Para el ejemplo del examen SAT presentado del ejercicio 2.16, haga lo siguiente:

*a*) Grafique la calificación del examen de lectura de mujeres contra la calificación del examen de lectura de hombres.

*b*) Si el diagrama de dispersión indica que parece apropiado establecer una relación lineal entre los dos, obtenga la regresión de la calificación del examen de lectura de mujeres sobre la calificación del examen de lectura de hombres.

*c*) De haber una relación entre las dos calificaciones de lectura, ¿dicha relación sería *causal*?

3.26. Repita el ejercicio 3.25, pero ahora sustituya las calificaciones de lectura con las de matemáticas.

3.27. *Trabajo para realizar en clase* sobre el estudio Monte Carlo: Consulte los 10 valores de $X$ de la tabla 3.2. Sea $\beta_1 = 25$ y $\beta_2 = 0.5$. Suponga que $u_i \approx N(0, 9)$, es decir, las $u_i$ están normalmente distribuidas con media 0 y varianza 9. Genere 100 muestras con estos valores para obtener 100 valores estimados de $\beta_1$ y $\beta_2$. Grafique estos valores estimados. ¿A qué conclusiones llega a partir del estudio Monte Carlo? *Nota*: ahora la mayor parte de los paquetes estadísticos genera variables aleatorias de casi todas las distribuciones de probabilidad conocidas. Pida ayuda a su profesor en caso de que tenga dificultad para generar dichas variables.

3.28 Con los datos de la tabla 3.3, obtenga la gráfica del número de suscriptores de teléfonos celulares en relación con el número de computadoras personales en uso. ¿Hay alguna relación discernible entre los dos? De ser así, ¿cómo la explicaría?

# Apéndice 3A

## 3A.1 Derivación de estimados de mínimos cuadrados

Al diferenciar parcialmente (3.1.2) respecto de $\hat{\beta}_1$ y $\hat{\beta}_2$, se obtiene

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_1} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum \hat{u}_i \tag{1}$$

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_2} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i = -2\sum \hat{u}_i X_i \tag{2}$$

Se igualan estas ecuaciones a cero y, después de alguna simplificación y manipulación algebraica, se obtienen los estimadores dados en las ecuaciones (3.1.6) y (3.1.7).

## 3A.2 Propiedades de linealidad e insesgamiento de los estimadores de mínimos cuadrados

De (3.1.8) se tiene

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i \tag{3}$$

donde

$$k_i = \frac{x_i}{\left(\sum x_i^2\right)}$$

lo cual muestra que $\hat{\beta}_2$ es un **estimador lineal** porque es una función lineal de $Y$; de hecho, es un promedio ponderado de $Y_i$ donde $k_i$ representa las ponderaciones. De la misma manera se demuestra que $\hat{\beta}_1$ también es un estimador lineal.

Por cierto, note las siguientes propiedades de las ponderaciones $k_i$:

1. Como se supuso que las $X_i$ son no estocásticas, las $k_i$ también son no estocásticas
2. $\sum k_i = 0$.
3. $\sum k_i^2 = 1\big/\sum x_i^2$.
4. $\sum k_i x_i = \sum k_i X_i = 1$. Estas propiedades se verifican directamente con la definición de $k_i$.

Por ejemplo,

$$\sum k_i = \sum\left(\frac{x_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2}\sum x_i, \quad \text{pues, para una muestra dada, se conoce } \sum x_i^2$$

$$= 0, \quad \text{pues } \sum x_i\text{, la suma de las desviaciones de la media, es siempre cero.}$$

Ahora sustituya la FRP $Y_i = \beta_1 + \beta_2 X_i + u_i$ en la ecuación (3) para obtener

$$\begin{aligned}\hat{\beta}_2 &= \sum k_i(\beta_1 + \beta_2 X_i + u_i)\\ &= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i\\ &= \beta_2 + \sum k_i u_i\end{aligned} \tag{4}$$

donde se emplean las propiedades de $k_i$ anotadas antes.

Ahora, al obtener los valores esperados de la ecuación (4) para ambos lados y advertir que las $k_i$, al ser no estocásticas, pueden tratarse como constantes, se obtiene:

$$\begin{aligned} E(\hat{\beta}_2) &= \beta_2 + \sum k_i E(u_i) \\ &= \beta_2 \end{aligned} \tag{5}$$

pues $E(u_i) = 0$ por suposición. Por consiguiente, $\hat{\beta}_2$ es un estimador insesgado de $\beta_2$. De la misma manera se demuestra que $\hat{\beta}_1$ es también un estimador insesgado de $\beta_1$.

## 3A.3 Varianzas y errores estándar de los estimadores de mínimos cuadrados

Ahora, de acuerdo con la definición de varianza, se escribe

$$\begin{aligned} \text{var}(\hat{\beta}_2) &= E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2 \\ &= E(\hat{\beta}_2 - \beta_2)^2 \qquad \text{porque } E(\hat{\beta}_2) = \beta_2 \\ &= E\left(\sum k_i u_i\right)^2 \qquad \text{con la ecuación (4) anterior} \\ &= E\left(k_1^2 u_1^2 + k_2^2 u_2^2 + \cdots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + \cdots + 2k_{n-1} k_n u_{n-1} u_n\right) \end{aligned} \tag{6}$$

Por los supuestos $E(u_i^2) = \sigma^2$ para cada $i$ y $E(u_i u_j) = 0$, $i \neq j$, se deduce que

$$\begin{aligned} \text{var}(\hat{\beta}_2) &= \sigma^2 \sum k_i^2 \\ &= \frac{\sigma^2}{\sum x_i^2} \qquad \text{(con la definición de } k_i^2\text{)} \\ &= \text{Ecuación (3.3.1)} \end{aligned} \tag{7}$$

La varianza de $\hat{\beta}_1$ se obtiene con el mismo razonamiento. Una vez obtenidas las varianzas de $\hat{\beta}_1$ y $\hat{\beta}_2$, se obtienen también los errores estándar correspondientes, al tomar las raíces cuadradas positivas.

## 3A.4 Covarianza entre $\hat{\beta}_1$ y $\hat{\beta}_2$

Por definición,

$$\begin{aligned} \text{cov}(\hat{\beta}_1, \hat{\beta}_2) &= E\{[\hat{\beta}_1 - E(\hat{\beta}_1)][\hat{\beta}_2 - E(\hat{\beta}_2)]\} \\ &= E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \qquad \text{(¿Por qué?)} \\ &= -\bar{X} E(\hat{\beta}_2 - \beta_2)^2 \\ &= -\bar{X}\, \text{var}(\hat{\beta}_2) \\ &= \text{Ecuación (3.3.9)} \end{aligned} \tag{8}$$

donde se aprovecha que $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$ y $E(\hat{\beta}_1) = \bar{Y} - \beta_2 \bar{X}$, que es igual a $\hat{\beta}_1 - E(\hat{\beta}_1) = -\bar{X}(\hat{\beta}_2 - \beta_2)$. *Nota*: la var($\hat{\beta}_2$) está dada en la ecuación (3.3.1).

## 3A.5 Estimador de mínimos cuadrados de $\sigma^2$

Recuerde que

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{9}$$

Por consiguiente,

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u} \tag{10}$$

Al restar la ecuación (10) de la ecuación (9) se obtiene

$$y_i = \beta_2 x_i + (u_i - \bar{u}) \tag{11}$$

Recuerde también que

$$\hat{u}_i = y_i - \hat{\beta}_2 x_i \tag{12}$$

Por tanto, al sustituir (11) en (12) se obtiene

$$\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i \tag{13}$$

Se reúnen términos, se eleva al cuadrado y se suman ambos lados para obtener:

$$\sum \hat{u}_i^2 = (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + \sum (u_i - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u}) \tag{14}$$

Al tomar valores esperados en ambos lados se tiene

$$\begin{aligned} E\left(\sum \hat{u}_i^2\right) &= \sum x_i^2 E(\hat{\beta}_2 - \beta_2)^2 + E\left[\sum (u_i - \bar{u})^2\right] - 2E\left[(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u})\right] \\ &= \sum x_i^2 \operatorname{var}(\hat{\beta}_2) + (n-1) \operatorname{var}(u_i) - 2E\left[\sum k_i u_i (x_i u_i)\right] \\ &= \sigma^2 + (n-1)\sigma^2 - 2E\left[\sum k_i x_i u_i^2\right] \\ &= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 \\ &= (n-2)\sigma^2 \end{aligned} \tag{15}$$

donde, en el penúltimo paso, se utiliza la definición de $k_i$ dada en la ecuación (3) y la relación dada en la ecuación (4). También observe que

$$\begin{aligned} E \sum (u_i - \bar{u})^2 &= E\left[\sum u_i^2 - n\bar{u}^2\right] \\ &= E\left[\sum u_i^2 - n\left(\frac{\sum u_i}{n}\right)^2\right] \\ &= E\left[\sum u_i^2 - \frac{1}{n} \sum \left(u_i^2\right)\right] \\ &= n\sigma^2 - \frac{n}{n}\sigma^2 = (n-1)\sigma^2 \end{aligned}$$

donde se aprovecha que las $u_i$ están no correlacionadas y que la varianza de cada $u_i$ es $\sigma^2$.

Así, se obtiene

$$E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2 \tag{16}$$

Por tanto, si se define

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{17}$$

su valor esperado es

$$E(\hat{\sigma}^2) = \frac{1}{n-2} E\left(\sum \hat{u}_i^2\right) = \sigma^2 \qquad \text{con la ecuación (16)} \tag{18}$$

lo cual muestra que $\hat{\sigma}^2$ es un estimador insesgado de la verdadera $\sigma^2$.

## 3A.6 Propiedad de varianza mínima de los estimadores de mínimos cuadrados

Se demostró en el apéndice 3A, sección 3A.2, que el estimador $\hat{\beta}_2$ de mínimos cuadrados $\hat{\beta}_2$ era lineal e insesgado (esto vale también para $\hat{\beta}_1$). Para demostrar que estos estimadores tienen varianza mínima dentro de la clase de todos los estimadores lineales insesgados, considere el estimador de mínimos cuadrados $\hat{\beta}_2$:

$$\hat{\beta}_2 = \sum k_i Y_i$$

donde

$$k_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} = \frac{x_i}{\sum x_i^2} \qquad \text{(véase el apéndice 3A.2)} \tag{19}$$

lo cual demuestra que $\hat{\beta}_2$ es un promedio ponderado de las *Y*, con las $k_i$ como ponderaciones.

Se define un estimador lineal alterno de $\beta_2$ de la siguiente forma:

$$\beta_2^* = \sum w_i Y_i \tag{20}$$

donde $w_i$ son también ponderaciones, no necesariamente iguales a $k_i$. Ahora

$$\begin{aligned} E(\beta_2^*) &= \sum w_i E(Y_i) \\ &= \sum w_i (\beta_1 + \beta_2 X_i) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i X_i \end{aligned} \tag{21}$$

Por consiguiente, para que ($\beta_2^*$) sea insesgado se requiere que

$$\sum w_i = 0 \tag{22}$$

y

$$\sum w_i X_i = 1 \tag{23}$$

También se puede escribir

$$\begin{aligned} \operatorname{var}(\beta_2^*) &= \operatorname{var} \sum w_i Y_i \\ &= \sum w_i^2 \operatorname{var} Y_i \qquad [\textit{Nota:} \operatorname{var} Y_i = \operatorname{var} u_i = \sigma^2] \\ &= \sigma^2 \sum w_i^2 \qquad [\textit{Nota:} \operatorname{cov}(Y_i, Y_j) = 0\,(i \neq j)] \\ &= \sigma^2 \sum \left( w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2 \qquad \text{(Observe la manipulación matemática)} \\ &= \sigma^2 \sum \left( w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \frac{\sum x_i^2}{\left(\sum x_i^2\right)^2} + 2\sigma^2 \sum \left( w_i - \frac{x_i}{\sum x_i^2} \right) \left( \frac{x_i}{\sum x_i^2} \right) \\ &= \sigma^2 \sum \left( w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left( \frac{1}{\sum x_i^2} \right) \end{aligned} \tag{24}$$

porque el último término desaparece en el penúltimo paso. (¿Por qué?)

Como el último término en la ecuación (24) es constante, la varianza de ($\beta_2^*$) sólo se reduce al manipular el primer término. Sea

$$w_i = \frac{x_i}{\sum x_i^2}$$

La ecuación (24) se reduce a

$$\begin{aligned} \operatorname{var}(\beta_2^*) &= \frac{\sigma^2}{\sum x_i^2} \\ &= \operatorname{var}(\hat{\beta}_2) \end{aligned} \tag{25}$$

Expresado en palabras, con ponderaciones $w_i = k_i$, que son ponderaciones de mínimos cuadrados, la varianza del estimador lineal $\beta_2^*$ es igual a la del estimador de mínimos cuadrados $\hat{\beta}_2$; de lo contrario, la $\text{var}(\beta_2^*) > \text{var}(\hat{\beta}_2)$. Dicho de otra manera, si hay un estimador lineal insesgado de $\beta_2$ de varianza mínima, debe ser el estimador de mínimos cuadrados. Del mismo modo se demuestra que $\hat{\beta}_1$ es un estimador lineal insesgado con varianza mínima de $\beta_1$.

## 3A.7 Consistencia de los estimadores de mínimos cuadrados

Ya se demostró que, en el marco teórico del modelo de regresión lineal clásica, los estimadores mínimos cuadrados son insesgados (y eficientes) con cualquier tamaño de muestra, grande o pequeña. Pero a veces, como se analiza en el **apéndice A**, un estimador no satisface una o más propiedades estadísticas deseables en pequeñas muestras. Sin embargo, a medida que el tamaño de la muestra crece indefinidamente, los estimadores poseen diversas propiedades estadísticas deseables. Tales propiedades se conocen como **propiedades asintóticas** o **de muestras grandes**. En el apéndice se analizará una propiedad de muestra grande, a saber, la propiedad de **consistencia**, que se estudia con mayor detalle en el **apéndice A**. Para el modelo de dos variables ya se mostró que el estimador de MCO $\hat{\beta}_2$ es un estimador insesgado del verdadero $\beta_2$. Ahora se demuestra que $\hat{\beta}_2$ es también un estimador consistente de $\beta_2$. Como se analiza en el **apéndice A**, una condición suficiente para la consistencia es que $\hat{\beta}_2$ sea insesgado y que su varianza tienda a cero conforme el tamaño de la muestra $n$ tienda al infinito.

Como ya se probó la propiedad de insesgamiento, sólo se necesita demostrar que la varianza de $\hat{\beta}_2$ tiende a cero conforme $n$ crece indefinidamente. Se sabe que

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2/n}{\sum x_i^2/n} \tag{26}$$

Al dividir el numerador y el denominador por $n$, no se modifica la igualdad.

Ahora

$$\underbrace{\text{lím var}(\hat{\beta}_2)}_{n\to\infty} = \underbrace{\text{lím}\left(\frac{\sigma^2/n}{\sum x_i^2/n}\right)}_{n\to\infty} = 0 \tag{27}$$

donde se aprovecha que 1) el límite de una razón es el límite del numerador entre el límite del denominador (consulte un libro de cálculo); 2) conforme $n$ tiende a infinito, $\sigma^2/n$ tiende a cero porque $\sigma^2$ es un número finito, y $[(\sum x_i^2)/n] \neq 0$ porque la varianza de $X$ tiene un límite finito en vista del supuesto 7 del MCRL.

La conclusión del análisis anterior es que el estimador de MCO $\hat{\beta}_2$ es un estimador consistente del verdadero $\beta_2$. De igual forma, se establece que $\hat{\beta}_1$ también es un estimador consistente. Por tanto, en repetidas muestras (pequeñas), los estimadores de MCO son insesgados, y conforme la muestra aumenta indefinidamente, los estimadores de MCO son consistentes. Como se verá más tarde, aunque no se satisfagan algunas suposiciones del MCRL, en diversas situaciones es posible obtener estimadores consistentes de los coeficientes de regresión.

Capítulo 4

# Modelo clásico de regresión lineal normal (MCRLN)

La llamada **teoría clásica de la inferencia estadística** consta de dos ramas, a saber: **estimación** y **pruebas de hipótesis**. Hasta el momento hemos estudiado el tema de la estimación de los parámetros del modelo de regresión lineal (con dos variables). Mediante el método de MCO fue posible estimar los parámetros $\beta_1$, $\beta_2$ y $\sigma^2$. Con los supuestos del *modelo clásico de regresión lineal* (MCRL) demostramos que los estimadores de dichos parámetros, $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\sigma}^2$, satisfacen varias propiedades estadísticas deseables, como el insesgamiento, la varianza mínima, etc. (Recuerde la propiedad MELI.) Observe que, en vista de que son estimadores, sus valores cambiarán de muestra en muestra. Por consiguiente, tales estimadores son *variables aleatorias*.

Pero la estimación es sólo la mitad de la batalla. Las pruebas de hipótesis constituyen la otra mitad. Tenga presente que, en el análisis de regresión, nuestro objetivo no sólo consiste en estimar la función de regresión muestral (FRM), sino también en utilizarla para obtener inferencias respecto de la función de regresión poblacional (FRP), como destacamos en el capítulo 2. Así, es conveniente saber qué tan cerca está $\hat{\beta}_1$ del verdadero valor de $\beta_1$, o qué tan cerca está $\hat{\sigma}^2$ del verdadero $\sigma^2$. Por ejemplo, en el ejemplo 3.2 estimamos la FRM como se muestra en la ecuación (3.7.2). Pero, en vista de que la regresión se basa en una muestra de 55 familias, ¿cómo saber si la PMC estimada de 0.4368 representa la PMC (verdadera) en la población total?

Por tanto, como $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\sigma}^2$ son variables aleatorias, es necesario averiguar sus distribuciones de probabilidad, pues sin conocerlas no es posible relacionarlas con sus valores verdaderos.

## 4.1 Distribución de probabilidad de las perturbaciones $u_i$

Para encontrar las distribuciones de probabilidad de los estimadores de MCO, procederemos como sigue. De manera específica, consideramos $\hat{\beta}_2$. Como vimos en el apéndice 3A.2,

$$\hat{\beta}_2 = \sum k_i Y_i \tag{4.1.1}$$

donde $k_i = x_i / \sum x_i^2$. Sin embargo, como se supone que las $X$ son fijas, o no estocásticas, debido a que se trata de un análisis de regresión condicional (condicional en los valores fijos de $X_i$), la

ecuación (4.1.1) muestra que $\hat{\beta}_2$ es una función *lineal* de $Y_i$, la cual se supone aleatoria. Pero, como $Y_i = \beta_1 + \beta_2 X_i + u_i$, expresamos (4.1.1) como

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) \qquad \textbf{(4.1.2)}$$

Debido a que $k_i$, las betas, y $X_i$ son fijas, $\hat{\beta}_2$ es a fin de cuentas una función *lineal* de la variable aleatoria $u_i$, aleatoria por suposición. Por tanto, la distribución de probabilidad de $\hat{\beta}_2$ (y también de $\hat{\beta}_1$) dependerá del supuesto que se hizo respecto de la distribución de probabilidad de $u_i$. Y como se requiere conocer las distribuciones de probabilidad de los estimadores de MCO para obtener las inferencias sobre sus valores poblacionales, la naturaleza de la distribución de probabilidad de $u_i$ desempeña un papel importante en las pruebas de hipótesis.

Debido a que el método de MCO no hace ninguna suposición respecto de la naturaleza probabilística de $u_i$, resulta de poca ayuda para el propósito de hacer inferencias sobre la FRP mediante la FRM, a pesar del teorema de Gauss-Markov. Este vacío puede llenarse si se supone que las $u$ siguen una determinada distribución de probabilidad. Por razones que mencionaremos en seguida, en el contexto de regresión se supone, por lo general, que las $u$ tienen la distribución de probabilidad normal. Si a los supuestos del modelo clásico de regresión lineal (MCRL) analizados en el capítulo 3 se añade el supuesto de normalidad para $u_i$, obtenemos lo que se conoce como **modelo clásico de regresión lineal normal (MCRLN)**.

## 4.2 Supuesto de normalidad de $u_i$

El modelo clásico de regresión lineal normal supone que cada $u_i$ está *normalmente* distribuida con

Media: $$E(u_i) = 0 \qquad \textbf{(4.2.1)}$$

Varianza: $$E[u_i - E(u_i)]^2 = E\left(u_i^2\right) = \sigma^2 \qquad \textbf{(4.2.2)}$$

cov($u_i$, $u_j$): $$E\{[(u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i\, u_j) = 0 \quad i \neq j \qquad \textbf{(4.2.3)}$$

Estos supuestos se expresan en forma más compacta como

$$u_i \sim N(0, \sigma^2) \qquad \textbf{(4.2.4)}$$

donde el símbolo $\sim$ significa *distribuido* y $N$ significa *distribución normal,* y donde los términos entre paréntesis representan los dos parámetros de la distribución normal: la media y la varianza.

Como se señala en el **apéndice A,** para **dos variables normalmente distribuidas, una covarianza o correlación cero significa independencia entre las dos variables**. Por consiguiente, con el supuesto de normalidad, la ecuación (4.2.4) significa que $u_i$ y $u_j$ no sólo no están correlacionadas, sino que también están independientemente distribuidas.

Por tanto, (4.2.4) se escribe como

$$u_i \sim \text{NID}(0, \sigma^2) \qquad \textbf{(4.2.5)}$$

donde **NID** significa *normal e independientemente distribuido*.

## ¿Por qué debe formularse el supuesto de normalidad?

¿Por qué se emplea el supuesto de normalidad? Existen diversas razones.

1. Como señalamos en la sección 2.5, $u_i$ representa la influencia combinada (sobre la variable dependiente) de un gran número de variables independientes que no se introdujeron explícitamente en el modelo de regresión. Como explicamos, se espera que la influencia de estas variables omitidas o descartadas sea pequeña y, en el mejor de los casos, aleatoria. Ahora, gracias al conocido **teorema central del límite (TCL)** en estadística (véanse los detalles en el **apéndice A**), se puede demostrar que, si existe un gran número de variables aleatorias independientes con idéntica distribución, entonces, con pocas excepciones, la distribución de su suma tiende a ser normal a medida que se incrementa al infinito el número de tales variables.[1] Este teorema del límite central es el que proporciona una justificación teórica para el supuesto de normalidad de $u_i$.

2. Una variante del teorema del límite central establece que, aunque el número de variables no sea muy grande, o si estas variables no son estrictamente independientes, su suma puede estar aún normalmente distribuida.[2]

3. Con el supuesto de normalidad, se derivan con facilidad las distribuciones de probabilidad de los estimadores de MCO, pues, como se explica en el **apéndice A**, una propiedad de la distribución normal es que **cualquier función lineal de variables normalmente distribuidas estará también normalmente distribuida**. Como ya analizamos, los estimadores de MCO $\hat{\beta}_1$ y $\hat{\beta}_2$ son funciones lineales de $u_i$. Por consiguiente, si $u_i$ está normalmente distribuida, también lo están $\hat{\beta}_1$ y $\hat{\beta}_2$, lo cual hace que la tarea de probar hipótesis sea muy fácil.

4. La distribución normal es una distribución comparativamente sencilla y requiere sólo dos parámetros (la media y la varianza); es muy conocida y sus propiedades teóricas se han estudiado con amplitud en estadística matemática. Además, al parecer muchos fenómenos se rigen por la distribución normal.

5. Si trabajamos con una muestra finita o pequeña, con datos de 100 o menos observaciones, la suposición de normalidad desempeña un papel relevante. No sólo contribuye a derivar las distribuciones de probabilidad exactas de los estimadores de MCO, sino también permite utilizar las pruebas estadísticas $t$, $F$ y $\chi^2$ para los modelos de regresión. Las propiedades estadísticas de las distribuciones estadísticas $t$, $F$ y $\chi^2$ se estudian en el **apéndice A**. Como veremos en seguida, si el tamaño de la muestra es razonablemente grande, se puede flexibilizar el supuesto de normalidad.

6. Por último, en *muestras grandes,* los estadísticos $t$ y $F$ tienen aproximadamente las distribuciones de probabilidad de $t$ y $F$, por lo que las pruebas $t$ y $F$ que se basan en el supuesto de que el término de error está distribuido normalmente pueden seguir aplicándose con validez.[3] En la actualidad hay muchos datos transversales y de series de tiempo con una cantidad relativamente grande de observaciones. Por tanto, el supuesto de normalidad puede no ser tan crucial en conjuntos grandes de datos.

*Advertencia:* Como se está "imponiendo" el supuesto de normalidad, es menester encontrar aplicaciones prácticas que requieran tamaños pequeños de muestras en las que el supuesto de normalidad resulte apropiado. Más adelante se realizarán algunas pruebas para hacer precisa-

---

[1] Para un análisis relativamente sencillo y directo de este teorema, véase Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 2a. ed., Harcourt Academic Press, Nueva York, 2000, pp. 193-194. Una excepción a este teorema es la distribución de Cauchy, la cual no tiene media ni momentos más altos. Véase M. G. Kendall y A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin & Co., Londres, 1960, vol. 1, pp. 248-249.

[2] Para las diversas formas del teorema central del límite, véase Harald Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, Nueva Jersey, 1946, cap. 17.

[3] Para un análisis técnico de este punto, véase Christiaan Heij *et al.*, *Econometric Methods with Applications in Business and Economics*, Oxford University Press, Oxford, 2004, p. 197.

mente eso; asimismo, se presentarán situaciones en las que tal vez sea inadecuado el supuesto de normalidad. No obstante, hasta ese momento, consideraremos válido el supuesto de normalidad por las razones expuestas.

## 4.3 Propiedades de los estimadores de MCO según el supuesto de normalidad

Si suponemos que $u_i$ sigue la distribución normal, como en (4.2.5), los estimadores de MCO tienen las propiedades que se mencionan a continuación (el **apéndice A** ofrece un análisis general de las propiedades estadísticas deseables de los estimadores):

1. Son insesgados.

2. Tienen varianza mínima. En combinación con 1, esto significa que son **estimadores insesgados con varianza mínima**, o **eficientes**.

3. Presentan **consistencia**; es decir, a medida que el tamaño de la muestra aumenta indefinidamente, los estimadores convergen hacia sus verdaderos valores poblacionales.

4. $\hat{\beta}_1$ (al ser una función lineal de $u_i$) está *normalmente distribuida* con

$$\text{Media:} \quad E(\hat{\beta}_1) = \beta_1 \tag{4.3.1}$$

$$\text{var}(\hat{\beta}_1): \quad \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \qquad = (3.3.3) \tag{4.3.2}$$

O, en forma más compacta,

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$$

Entonces, de acuerdo con las propiedades de la distribución normal, la variable $Z$, definida como

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \tag{4.3.3}$$

sigue la **distribución normal estándar**, es decir, una distribución normal con media cero y varianza unitaria (= 1), o

$$Z \sim N(0, 1)$$

5. $\hat{\beta}_2$ (al ser una función lineal de $u_i$) está *normalmente* distribuida con

$$\text{Media:} \quad E(\hat{\beta}_2) = \beta_2 \tag{4.3.4}$$

$$\text{var}(\hat{\beta}_2): \quad \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2} \qquad = (3.3.1) \tag{4.3.5}$$

O, en forma más compacta,

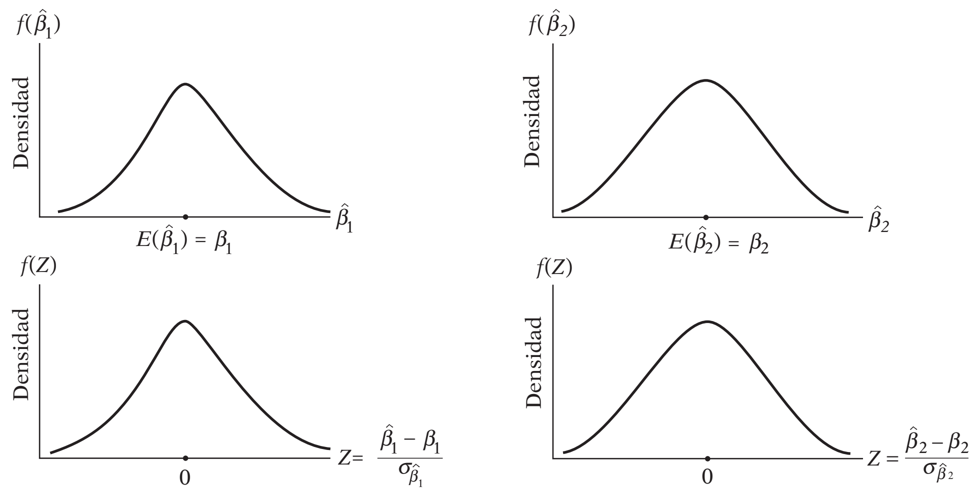
$$\hat{\beta}_2 \sim N\left(\beta_2, \sigma^2_{\hat{\beta}_2}\right)$$

Entonces, como en (4.3.3),

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \tag{4.3.6}$$

también sigue una distribución normal estándar.

**FIGURA 4.1**
Distribuciones de probabilidad de $\hat{\beta}_1$ y $\hat{\beta}_2$.




Geométricamente, las distribuciones de probabilidad de $\hat{\beta}_1$ y $\hat{\beta}_2$ se muestran en la figura 4.1.

6. $(n - 2)(\hat{\sigma}^2/\sigma^2)$ está distribuida como la distribución $\chi^2$ (ji cuadrada), con $(n - 2)$ gl.[4] Saber esto ayuda a hacer inferencias respecto de la verdadera $\sigma^2$ a partir de la $\sigma^2$ estimada, como veremos en el capítulo 5. (En el **apéndice A** se analizan la distribución ji cuadrada y sus propiedades.)

7. $(\hat{\beta}_1, \hat{\beta}_2)$ se distribuyen de manera independiente respecto de $\hat{\sigma}^2$. La importancia de lo anterior se explica en el siguiente capítulo.

8. $\hat{\beta}_1$ *y* $\hat{\beta}_2$ *tienen varianza mínima entre todas las clases de estimadores insesgados, lineales o no lineales*. Este resultado, desarrollado por Rao, es muy eficaz porque, a diferencia del teorema de Gauss-Markov, no se limita a la clase de estimadores lineales.[5] Por tanto, se puede decir que los estimadores de mínimos cuadrados son los **mejores estimadores insesgados (MEI)**; es decir, tienen varianza mínima en toda la clase de los estimadores insesgados.

**Para resumir**: lo importante es que el supuesto de normalidad permite derivar las distribuciones de probabilidad, o muestrales, de $\hat{\beta}_1$ y $\hat{\beta}_2$ (ambas normales), y de $\hat{\sigma}^2$ (relacionada con ji cuadrada). Como veremos en el capítulo 5, esto simplifica la tarea de establecer intervalos de confianza y pruebas (estadísticas) de hipótesis.

A propósito, observe que si supone que $u_i \sim N(0, \sigma^2)$, $Y_i$, al ser una función lineal de $u_i$, posee también una distribución normal con una media y una varianza dadas por

$$E(Y_i) = \beta_1 + \beta_2 X_i \tag{4.3.7}$$

$$\text{var}(Y_i) = \sigma^2 \tag{4.3.8}$$

En forma más ordenada,

$$Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2) \tag{4.3.9}$$

---

[4] La prueba de este enunciado es un poco intrincada. Se proporciona una fuente accesible para la demostración en la obra de Robert V. Hogg y Allen T. Craig, *Introduction to Mathematical Statistics*, 2a. ed., Macmillan, Nueva York, 1965, p. 144.

[5] C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Nueva York, 1965, p. 258.

## 4.4 Método de máxima verosimilitud (MV)

Un método de estimación puntual con algunas propiedades teóricamente más fuertes que las del método de MCO es el método de **máxima verosimilitud (MV)**. Como este método es un poco complicado, se analiza en el apéndice de este capítulo. Para el lector que sólo tiene un interés general, baste aclarar que si se supuso que $u_i$ está normalmente distribuida, como lo hemos hecho por las razones expuestas, los estimadores de MV y MCO de los coeficientes de regresión, los $\beta$, son idénticos, y esto es válido para regresiones simples al igual que para las regresiones múltiples. El estimador de MV de $\sigma^2$ es $\sum \hat{u}_i^2/n$. Este estimador es sesgado, en tanto que el estimador de MCO de $\sigma^2 = \sum \hat{u}_i^2/(n-2)$, como vimos, es insesgado. Pero, al comparar estos dos estimadores de $\sigma^2$, se ve que, conforme aumenta el tamaño de la muestra $n$, los dos estimadores de $\sigma^2$ tienden a ser iguales. Por tanto, asintóticamente (es decir, a medida que $n$ crece indefinidamente), el estimador MV de $\sigma^2$ también es insesgado.

Como el método de mínimos cuadrados con el supuesto adicional de normalidad de $u_i$ proporciona todas las herramientas necesarias para llevar a cabo la estimación y las pruebas de hipótesis de los modelos de regresión lineal, no existe pérdida alguna para los lectores que no deseen continuar revisando el método de máxima verosimilitud debido a su cierta complejidad matemática.

## Resumen y conclusiones

1. En este capítulo analizamos el modelo clásico de regresión lineal *normal* (MCRLN).
2. Este modelo difiere del modelo clásico de regresión lineal (MCRL) al suponer específicamente que el término de perturbación, $u_i$, que forma parte del modelo de regresión, está normalmente distribuido. El MCRL no requiere ningún supuesto sobre la distribución de probabilidad de $u_i$; sólo requiere que el valor de la media de $u_i$ sea cero y su varianza sea una constante finita.
3. La justificación teórica del supuesto de normalidad es el **teorema central del límite**.
4. Sin el supuesto de normalidad, según los otros supuestos analizados en el capítulo 3, el teorema de Gauss-Markov demostró que los estimadores de MCO son MELI.
5. Con el supuesto adicional de normalidad, los estimadores de MCO no sólo son los **mejores estimadores insesgados (MEI)** sino que también siguen distribuciones de probabilidad bien conocidas. Los estimadores de MCO del intercepto y de la pendiente están normalmente distribuidos, y el estimador de MCO de la varianza de $u_i$ $(= \hat{\sigma}^2)$ se relaciona con la distribución ji cuadrada.
6. En los capítulos 5 y 8 veremos la utilidad de estos conocimientos para realizar inferencias respecto de los valores de los parámetros poblacionales.
7. Una alternativa al método de mínimos cuadrados es el **método de máxima verosimilitud (MV)**. Sin embargo, para utilizar este método se debe hacer un supuesto sobre la distribución de probabilidad del término de perturbación $u_i$. En el contexto de regresión, el supuesto más común es que las $u_i$ siguen la distribución normal.
8. Según el supuesto de normalidad, los estimadores de MCO y MV de los parámetros del intercepto y de la pendiente del modelo de regresión son idénticos. Sin embargo, los estimadores de MCO y MV de la varianza de $u_i$ son diferentes. No obstante, en muestras grandes, estos dos estimadores convergen.
9. Por tanto, el método de MV suele recibir el nombre de *método de muestras grandes*. El método MV tiene una aplicación más extensa, pues se aplica también a modelos de regresión no lineal en los parámetros. En este último caso, MCO generalmente no se utiliza. Para mayor información sobre esto, véase el capítulo 14.
10. En este texto recurriremos en gran medida al método de MCO por razones prácticas: *a*) en comparación con el de MV, el de MCO es fácil de aplicar; *b*) los estimadores de MV y MCO de $\beta_1$ y $\beta_2$ son idénticos (lo cual se cumple también en regresión múltiple), y *c*) aun en muestras moderadamente grandes, los estimadores de MCO y MV de $\sigma^2$ no difieren considerablemente.

Sin embargo, para satisfacer al lector con formación matemática, se presenta una breve introducción al método de MV en el apéndice de este capítulo y también en el **apéndice A**.

# Apéndice 4A

## 4A.1 Estimación de máxima verosimilitud del modelo de regresión con dos variables

Supongamos que en el modelo con dos variables $Y_i = \beta_1 + \beta_2 X_i + u_i$, las $Y_i$ son independientes y normalmente distribuidas con media $= \beta_1 + \beta_2 X_i$ y varianza $= \sigma^2$. [Véase la ecuación (4.3.9).] Como resultado, la función de densidad de probabilidad conjunta de $Y_1, Y_2, \ldots, Y_n$, dadas las medias y varianzas anteriores, se escribe de la siguiente forma:

$$f(Y_1, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2)$$

Pero dada la independencia de las $Y$, esta función de densidad de probabilidad conjunta se escribe como el producto de las $n$ funciones de densidad individuales como

$$\begin{aligned} & f(Y_1, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \\ & \quad = f(Y_1 \mid \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 \mid \beta_1 + \beta_2 X_i, \sigma^2) \cdots f(Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \end{aligned} \tag{1}$$

donde

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{2}$$

que es la función de densidad de una variable normalmente distribuida con media y varianza dadas.

(*Nota*: exp significa $e$ elevado a la potencia de la expresión indicada por $\{\}$.)

Al sustituir la ecuación (2) por cada $Y_i$ en (1) se tiene

$$f(Y_i, Y_2, \ldots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{3}$$

Si se conocen o están dadas $Y_1, Y_2, \ldots, Y_n$, pero no se conocen $\beta_1$, $\beta_2$ y $\sigma^2$, la función en (3) se llama **función de verosimilitud**, denotada con $\text{FV}(\beta_1, \beta_2, \sigma^2)$ y escrita como[1]

$$\text{FV}(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{4}$$

El **método de máxima verosimilitud**, como lo indica el nombre, consiste en estimar los parámetros desconocidos de manera que la probabilidad de observar las $Y$ dadas sea lo más alta (o máxima) posible. Por consiguiente, se tiene que encontrar el máximo de la función en la ecuación (4). Es un ejercicio sencillo de cálculo diferencial. Para la diferenciación, es más fácil expresar (4) en términos de la función logaritmo o log de la siguiente manera.[2] (*Nota*: ln = logaritmo natural.)

$$\begin{aligned} \ln \text{FV} &= -n \ln \sigma - \frac{n}{2} \ln (2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \\ &= -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln (2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \end{aligned} \tag{5}$$

[1] Por supuesto, si se conocen $\beta_1$, $\beta_2$ y $\sigma^2$ pero no las $Y_1$, la ecuación (4) representa la función de densidad de probabilidad conjunta: la probabilidad de observar conjuntamente las $Y_i$.

[2] Como la función log es una función monótona, ln FV alcanzará su máximo valor en el mismo punto que FV.

Al diferenciar (5) parcialmente respecto de $\beta_1$, $\beta_2$ y $\sigma^2$ se obtiene

$$\frac{\partial \ln \mathrm{FV}}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-1) \tag{6}$$

$$\frac{\partial \ln \mathrm{FV}}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-X_i) \tag{7}$$

$$\frac{\partial \ln \mathrm{FV}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_i)^2 \tag{8}$$

Igualamos estas ecuaciones a cero (la condición de primer orden para la optimización) y dejamos que $\tilde{\beta}_1$, $\tilde{\beta}_1$ y $\tilde{\sigma}^2$ denoten los estimadores de MV, para obtener[3]

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) = 0 \tag{9}$$

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)X_i = 0 \tag{10}$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = 0 \tag{11}$$

Después de simplificar, las ecuaciones (9) y (10) llevan a

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \tag{12}$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \tag{13}$$

las cuales son precisamente las *ecuaciones normales* de la teoría de mínimos cuadrados obtenidas en (3.1.4) y (3.1.5). Por tanto, los estimadores de MV, las $\tilde{\beta}$, son los mismos que los estimadores de MCO, los $\bar{\beta}$, dados en (3.1.6) y (3.1.7). Esta igualdad no es fortuita. Al examinar la verosimilitud (5) se ve que el último término entra con signo negativo. Por consiguiente, la maximización de (5) equivale a la minimización de este término, que es justo el enfoque de mínimos cuadrados, como se aprecia en (3.1.2).

Al sustituir los estimadores de MV (= MCO) en (11) y simplificar, se obtiene el estimador de MV de $\tilde{\sigma}^2$, así:

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 \\ &= \frac{1}{n}\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \\ &= \frac{1}{n}\sum \hat{u}_i^2\end{aligned} \tag{14}$$

Se deduce de (14) que el estimador de MV $\tilde{\sigma}^2$ difiere del estimador de MCO $\hat{\sigma}^2 = [1/(n-2)]\sum \hat{u}_i^2$, el cual, como se demuestra en el apéndice 3A, sección 3A.5, es un estimador insesgado de $\sigma^2$. Por tanto, el estimador de MV de $\sigma^2$ es sesgado. La magnitud de este sesgo se determina fácilmente de la siguiente manera:

---

[3] Se utiliza el símbolo ˜ (tilde) para los estimadores de MV y ˆ (acento circunflejo) para los estimadores de MCO.

Tomamos la esperanza matemática de (14) en ambos lados de la ecuación y se obtiene

$$
\begin{aligned}
E(\tilde{\sigma}^2) &= \frac{1}{n}E\left(\sum \hat{u}_i^2\right) \\
&= \left(\frac{n-2}{n}\right)\sigma^2 \qquad \text{con la ecuación (16) del apéndice 3A, sección 3A.5} \qquad (15)\\
&= \sigma^2 - \frac{2}{n}\sigma^2
\end{aligned}
$$

lo cual demuestra que $\tilde{\sigma}^2$ está sesgado hacia abajo (es decir, subestima el verdadero $\sigma^2$) en muestras pequeñas. Pero observe que a medida que se incrementa indefinidamente $n$, el tamaño de la muestra, el segundo término en (15), factor de sesgo, tiende a ser cero. Por consiguiente, *asintóticamente* (es decir, en una muestra muy grande), $\tilde{\sigma}^2$ también es *insesgada*. Es decir, el lím $E(\tilde{\sigma}^2) = \sigma^2$ a medida que $n \to \infty$. Se puede demostrar además que $\tilde{\sigma}^2$ es también un estimador **consistente**,[4] es decir, a medida que $n$ aumenta indefinidamente, $\tilde{\sigma}^2$ converge hacia su verdadero valor $\sigma^2$.

## 4A.2 Estimación de máxima verosimilitud del gasto en alimentos en India

Regresemos al ejemplo 3.2 y la ecuación (3.7.2), que representa la regresión del gasto dedicado a los alimentos del gasto total realizado por 55 familias rurales de India. Como al suponer la normalidad, los estimadores de MCO y ML de los coeficientes de regresión son iguales, los estimadores de MV resultan como $\tilde{\beta}_1 = \bar{\beta}_1 = 94.2087$ y $\tilde{\beta}_2 = \bar{\beta}_2 = 0.4368$. El estimador de MCO de $\sigma^2$ es $\hat{\sigma}^2 = 4\,469.6913$, pero el estimador de MV es $\tilde{\sigma}^2 = 4\,407.1563$, por lo que es más pequeño que el estimador de MCO. Como se observa, en muestras pequeñas el estimador de MV está sesgado hacia abajo; es decir, subestima en promedio la verdadera varianza $\sigma^2$. Por supuesto, como es de esperarse, conforme crezca la muestra, la diferencia entre los dos estimadores se reducirá. Al sustituir los valores de los estimadores en la función log de verosimilitud, resulta el valor de −308.1625. Si se desea maximizar el valor de FV, sólo se calcula el antilogaritmo de −308.1625. Ningún otro valor de los parámetros dará una probabilidad más alta de obtener la muestra que se utilizó en el análisis.

## Apéndice 4A Ejercicios

4.1. "Si dos variables aleatorias son estadísticamente independientes, el coeficiente de correlación entre las dos es cero. Pero lo contrario no necesariamente es cierto; es decir, una correlación de cero no implica independencia estadística. Sin embargo, cuando dos variables están normalmente distribuidas, una correlación cero necesariamente implica independencia estadística." Verifique esta afirmación para la siguiente función de densidad de probabilidad conjunta de dos variables normalmente distribuidas, $Y_1$ y $Y_2$ (esta función de densidad de probabilidad conjunta se conoce como **función de densidad de probabilidad normal bivariada**):

$$
\begin{aligned}
f(Y_1, Y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right. \\
&\quad \times \left.\left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{Y_2-\mu_2}{\sigma^2}\right)^2\right]\right\}
\end{aligned}
$$

[4] Véase el **apéndice A** para un análisis general de las propiedades de los estimadores de máxima verosimilitud, así como para la distinción entre insesgamiento asintótico y consistencia. En términos generales, en el insesgamiento asintótico se trata de encontrar el lím $E(\tilde{\sigma}_n^2)$ cuando $n$ tiende a infinito, donde $n$ es el tamaño de la muestra en la que se basa el estimador; en la consistencia, se trata de averiguar cómo se comporta $\tilde{\sigma}_n^2$ a medida que $n$ aumenta indefinidamente. Observe que la propiedad de insesgamiento es una propiedad del muestreo repetido de un estimador basado en una muestra de un tamaño dado, mientras que la consistencia se relaciona con el comportamiento del estimador a medida que aumenta indefinidamente el tamaño de la muestra.

donde: $\mu_1$ = media de $Y_1$
$\mu_2$ = media de $Y_2$
$\sigma_1$ = desviación estándar de $Y_1$
$\sigma_2$ = desviación estándar de $Y_2$
$\rho$ = coeficiente de correlación entre $Y_1$ y $Y_2$

4.2. Al aplicar las condiciones de segundo orden para la optimización (es decir, la prueba de la segunda derivada), demuestre que los estimadores de MV de $\beta_1$, $\beta_2$ y $\sigma^2$, que se obtienen al resolver las ecuaciones (9), (10) y (11), en realidad maximizan la función de verosimilitud (4).

4.3. Una variable aleatoria $X$ sigue la **distribución exponencial** si presenta la siguiente función de densidad de probabilidad (FDP):

$$f(X) = (1/\theta)e^{-X/\theta} \qquad \text{para } X > 0$$
$$= 0 \qquad \text{en otro caso}$$

donde $\theta > 0$ es el parámetro de la distribución. Con el método de MV demuestre que el estimador de MV de $\theta$ es $\hat{\theta} = \sum X_i/n$, donde $n$ es el tamaño de la muestra. Es decir, demuestre que el estimador de MV de $\theta$ es la media muestral $\bar{X}$.

4.4. Suponga que el resultado de un experimento se clasifica como un éxito o un fracaso. Sea $X = 1$ cuando el resultado es un éxito y $X = 0$ cuando es un fracaso; la densidad de probabilidad, o masa, función de $X$ está dada por

$$p(X = 0) = 1 - p$$
$$p(X = 1) = p, 0 \leq p \leq 1$$

¿Cuál es el estimador de máxima verosimilitud de $p$, la probabilidad de éxito?

Capítulo 5

# Regresión con dos variables: estimación por intervalos y pruebas de hipótesis

> Cuidado con comprobar demasiadas hipótesis: cuanto más se torturen los datos, más probable será que confiesen, pero la confesión obtenida bajo presión puede no ser admisible en el tribunal de la opinión científica.[1]

Como señalamos en el capítulo 4, la estimación y las pruebas de hipótesis constituyen las dos ramas principales de la estadística clásica. La teoría de la estimación consta de dos partes: estimación puntual y estimación por intervalos. En los dos capítulos anteriores estudiamos a fondo la estimación puntual, en donde se introdujeron los métodos MCO y MV de la estimación puntual. En este capítulo consideraremos primero la estimación por intervalos y luego trataremos el tema de las pruebas de hipótesis, muy relacionado con la estimación por intervalos.

## 5.1 Requisitos estadísticos

Antes de exponer el mecanismo preciso para la construcción de los intervalos de confianza y de las pruebas de hipótesis estadísticas, se supone que el lector está familiarizado con los conceptos fundamentales de probabilidad y estadística. Aunque el **apéndice A** no sustituye un curso básico de estadística, sí proporciona sus elementos esenciales con los cuales el lector deberá estar totalmente familiarizado. Conceptos importantes, como **probabilidad**, **distribuciones de probabilidad, errores tipo I y tipo II, nivel de significancia, potencia de una prueba estadística** e **intervalos de confianza** son cruciales para entender el material de este capítulo y los siguientes.

[1] Stephen M. Stigler, "Testing Hypothesis or Fitting Models? Another Look at Mass Extinctions", en Matthew H. Nitecki y Antoni Hoffman, eds., *Neutral Models in Biology*, Oxford University Press, Oxford, 1987, p. 148.

## 5.2 Estimación por intervalos: algunas ideas básicas

Para poner en orden las ideas, consideremos el ejemplo de los salarios y el nivel de escolaridad del capítulo 3. La ecuación (3.6.1) muestra que el incremento promedio estimado del salario medio por hora relacionado con un año de aumento en la escolaridad ($\hat{\beta}_2$) es de 0.7240, que constituye una cifra estimada (puntual) del valor poblacional desconocido $\beta_2$. ¿Qué tan confiable es esta estimación? Como mencionamos en el capítulo 3, debido a las fluctuaciones muestrales, es probable que una sola estimación difiera del valor verdadero, aunque en un muestreo repetido se espera que el promedio de los valores sea igual al valor verdadero. [*Nota:* $E(\hat{\beta}_2) = \beta_2$.] Ahora, en estadística, la confiabilidad de un estimador puntual se mide por su error estándar. Por tanto, en lugar de depender de un solo estimador puntual, se puede construir un intervalo alrededor del estimador puntual, por ejemplo, dentro de dos o tres errores estándar a cada lado del estimador puntual, tal que este intervalo tenga, por ejemplo, 95% de probabilidad de incluir al verdadero valor del parámetro. Ésta es, a grandes rasgos, la idea básica de la **estimación por intervalos**.

Para ser más específico, supongamos que se desea encontrar qué tan "cerca" está, por ejemplo, $\hat{\beta}_2$ de $\beta_2$. Con este fin, se trata de encontrar dos números positivos, $\delta$ y $\alpha$, este último situado entre 0 y 1, de modo que la probabilidad de que el **intervalo aleatorio** ($\hat{\beta}_2 - \delta$, $\hat{\beta}_2 + \delta$) contenga al verdadero $\beta_2$ sea $1 - \alpha$. Simbólicamente,

$$\Pr(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha \qquad \textbf{(5.2.1)}$$

Tal intervalo, si existe, se conoce como **intervalo de confianza**; a $1 - \alpha$ se le denomina **coeficiente de confianza**; y $\alpha$ $(0 < \alpha < 1)$ se conoce como **nivel de significancia**.[2] Los extremos del intervalo de confianza se conocen como **límites de confianza** (también denominados valores *críticos*), con $\hat{\beta}_2 - \delta$ como *límite* de **confianza inferior** y $\hat{\beta}_2 + \delta$ como *límite* de **confianza superior**. Observe que, en la práctica, $\alpha$ y $1 - \alpha$ suelen expresarse en forma porcentual como $100\alpha$ y $100(1 - \alpha)\%$.

La ecuación (5.2.1) muestra que un **estimador de intervalo**, en contraste con un estimador puntual, es un intervalo construido de manera que tenga una probabilidad específica $1 - \alpha$ de contener en sus límites al valor verdadero del parámetro. Por ejemplo, si $\alpha = 0.05$, o 5%, (5.2.1) debe leerse: la probabilidad de que el intervalo (aleatorio) que allí aparece incluya al verdadero $\beta_2$ es de 0.95, o 95%. El estimador por intervalo proporciona así una gama de valores dentro de los cuales puede encontrarse el verdadero $\beta_2$.

Es muy importante conocer los siguientes aspectos de la estimación por intervalos:

1. La ecuación (5.2.1) no afirma que la probabilidad de que $\beta_2$ se encuentre entre los límites dados sea $1 - \alpha$. Como se supone que $\beta_2$, aunque se desconoce, es un número fijo, se dice que está o no está dentro del intervalo. La ecuación (5.2.1) establece que, al utilizar el método descrito en este capítulo, la probabilidad de construir un intervalo que contenga $\beta_2$ es $1 - \alpha$.

2. El intervalo (5.2.1) es un **intervalo aleatorio**; es decir, variará de una muestra a la siguiente debido a que se basa en $\hat{\beta}_2$, el cual es aleatorio. (¿Por qué?)

3. Como el intervalo de confianza es aleatorio, los enunciados probabilísticos que le corresponden deben entenderse en un sentido de largo plazo, es decir, para muestreo repetido. Más

[2] También conocida como **probabilidad de cometer un error tipo I**. Un error tipo I consiste en rechazar una hipótesis verdadera, mientras que el error tipo II consiste en aceptar una hipótesis falsa. (Este tema se analiza con mayor detalle en el **apéndice A**.) El símbolo $\alpha$ se conoce también como **tamaño de la prueba (estadística)**.

específicamente, (5.2.1) significa: si se construyen muchos intervalos de confianza como el anterior con base probabilística de $1 - \alpha$, a la larga, en promedio, tales intervalos contendrán, en $1 - \alpha$ de los casos, el valor verdadero del parámetro.

4. Como mencionamos en 2, el intervalo (5.2.1) es aleatorio siempre y cuando $\hat{\beta}_2$ sea desconocido. Sin embargo, una vez que se tenga una muestra específica y se obtenga un valor numérico específico de $\hat{\beta}_2$, el intervalo (5.2.1) deja de ser aleatorio, y queda entonces fijo. En este caso, **no se puede** hacer la afirmación probabilística (5.2.1); es decir, no se puede afirmar que la probabilidad de que un intervalo *fijo* dado incluya al verdadero $\beta_2$ sea $1 - \alpha$. En esta situación, $\beta_2$ está en el intervalo fijo o fuera de él. Por consiguiente, la probabilidad será 1 o 0. Por tanto, en el ejemplo de salarios y nivel de escolaridad, si el intervalo de confianza a 95% se obtuviera como $(0.5700 \leq \beta_2 \leq 0.8780)$, como demostraremos en breve en la ecuación (5.3.9), **no se puede** afirmar que la probabilidad de que este intervalo incluya al verdadero $\beta_2$ sea de 95%. Esa probabilidad es 1 o 0.

¿Cómo se construyen los intervalos de confianza? De la exposición anterior se espera que si se conocen las **distribuciones muestrales o de probabilidad** de los estimadores, se puedan hacer afirmaciones sobre intervalos de confianza como (5.2.1). En el capítulo 4 vimos que, con el supuesto de normalidad de las perturbaciones $u_i$, los estimadores de MCO $\hat{\beta}_1$ y $\hat{\beta}_2$ están también normalmente distribuidos, y que el estimador de MCO, $\hat{\sigma}^2$, se relaciona con la distribución $\chi^2$ (ji cuadrada). Entonces, parece que la labor de construir intervalos de confianza es muy sencilla. ¡Y, de hecho, lo es!

## 5.3 Intervalos de confianza para los coeficientes de regresión $\beta_1$ y $\beta_2$

### Intervalo de confianza para $\beta_2$

En el capítulo 4, sección 4.3, demostramos que con el supuesto de normalidad de $u_i$, los estimadores de MCO $\hat{\beta}_1$ y $\hat{\beta}_2$ son en sí mismos normalmente distribuidos con medias y varianzas allí establecidas. Por consiguiente, por ejemplo, la variable

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \tag{5.3.1}$$

como se anotó en (4.3.6), es una variable normal estandarizada. Por tanto, parece que se puede utilizar la distribución normal para hacer afirmaciones probabilísticas sobre $\beta_2$, siempre que se conozca la verdadera varianza poblacional $\sigma^2$. Si se conoce $\sigma^2$, una propiedad importante de una variable normalmente distribuida con media $\mu$ y varianza $\sigma^2$ es que el área bajo la curva normal entre $\mu \pm \sigma$ es cercana a 68%, que entre $\mu \pm 2\sigma$ es alrededor de 95%, y que entre los límites $\mu \pm 3\sigma$ el área es cercana a 99.7%.

Pero pocas veces se conoce $\sigma^2$ y, en la práctica, está determinada por el estimador insesgado $\hat{\sigma}^2$. Si se reemplaza $\sigma$ por $\hat{\sigma}$, (5.3.1) puede escribirse así

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} = \frac{\text{Estimador} - \text{Parámetro}}{\text{Error estándar estimado del estimador}} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \tag{5.3.2}$$

donde ee ($\hat{\beta}_2$) se refiere ahora al error estándar estimado. Se demuestra (véase el apéndice 5A, sección 5A.2) que la variable $t$, así definida, sigue la distribución $t$ con $n - 2$ gl. [Note la diferencia entre (5.3.1) y (5.3.2).] Por consiguiente, en lugar de utilizar la distribución normal, se puede utilizar la distribución $t$ para construir un intervalo de confianza para $\beta_2$ de la siguiente forma:

$$\Pr\left(-t_{\alpha/2} \leq t \leq t_{\alpha/2}\right) = 1 - \alpha \tag{5.3.3}$$

donde el valor $t$ en el centro de esta doble desigualdad es el valor $t$ dado por (5.3.2), y donde $t_{\alpha/2}$ es el valor de la variable $t$ obtenida de la distribución $t$ para un nivel de significancia de $\alpha/2$ y $n - 2$ gl; a menudo se denomina el valor **crítico** $t$ a un nivel de significancia $\alpha/2$. Al sustituir (5.3.2) en (5.3.3) se obtiene

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2}{\text{ee}\,(\hat{\beta}_2)} \leq t_{\alpha/2}\right] = 1 - \alpha \tag{5.3.4}$$

Reorganizamos (5.3.4) y obtenemos

$$\Pr\left[\beta_2 - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2)\right] = 1 - \alpha \qquad (5.3.5)^3$$

La ecuación (5.3.5) proporciona un **intervalo de confianza** para $\beta_2$ de 100 $(1 - \alpha)$%, que se escribe en forma más compacta como

Intervalo de confianza para $\beta_2$ a 100$(1 - \alpha)$%:

$$\hat{\beta}_2 \pm t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \tag{5.3.6}$$

Mediante argumentación análoga y con (4.3.1) y (4.3.2), se escribe:

$$\Pr\left[\hat{\beta}_1 - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_1)\right] = 1 - \alpha \tag{5.3.7}$$

o, en forma más compacta,

Intervalo de confianza para $\beta_1$ a 100$(1 - \alpha)$%:

$$\hat{\beta}_1 \pm t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_1) \tag{5.3.8}$$

Observe un rasgo importante de los intervalos de confianza dados en (5.3.6) y (5.3.8): en ambos casos *la amplitud del intervalo de confianza es proporcional al error estándar del estimador*. Es decir, entre más grande sea el error estándar, más amplio será el intervalo de confianza. Expresado de otra forma, mientras más grande sea el error estándar del estimador, mayor será la incertidumbre de estimar el verdadero valor del parámetro desconocido. Así, el error estándar de un estimador suele describirse como una medida de la **precisión** del estimador (es decir, con qué precisión mide el estimador al verdadero valor poblacional).

[3] Algunos autores prefieren escribir (5.3.5) con los gl indicados explícitamente. Por tanto, escribirían

$$\Pr\left[\hat{\beta}_2 - t_{(n-2),\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{(n-2)\alpha/2}\,\text{ee}\,(\hat{\beta}_2)\right] = 1 - \alpha$$

Pero, por simplicidad, utilizaremos la notación hasta ahora expuesta; el contexto aclara los gl adecuados.

De vuelta al ejemplo de regresión del capítulo 3 (sección 3.6) del salario promedio por hora ($Y$) y el nivel de escolaridad ($X$), recuerde que en la tabla 3.2 descubrimos que $\hat{\beta}_2 = 0.7240$; ee $(\hat{\beta}_2) = 0.0700$. Como hay 13 observaciones, los grados de libertad (gl) son 11. Si suponemos que $\alpha = 5\%$, es decir, un coeficiente de confianza a 95%, entonces la tabla $t$ muestra que para 11 gl el valor **crítico** $t_{\alpha/2} = 2.201$. Al sustituir estos valores en (5.3.5), el lector debe verificar que el intervalo de confianza para $\beta_2$ a 95% sea el siguiente:[4]

$$0.5700 \leq \beta_2 \leq 0.8780 \tag{5.3.9}$$

O, con la ecuación (5.3.6), es

$$0.7240 \pm 2.201(0.0700)$$

es decir,

$$0.7240 \pm 0.1540 \tag{5.3.10}$$

**La interpretación de este intervalo de confianza es**: Dado el coeficiente de confianza de 95%, en 95 de cada 100 casos, los intervalos como la ecuación 5.3.9 contendrán al verdadero valor de $\beta_2$. Pero, como ya advertimos, no se puede afirmar que la probabilidad de que el intervalo específico de la ecuación (5.3.9) contenga al verdadero $\beta_2$ sea de 95%, porque este intervalo es ahora fijo y no aleatorio; por consiguiente, $\beta_2$ se encontrará o no dentro de él: la probabilidad de que el intervalo fijo específico incluya al verdadero valor de $\beta_2$ es por consiguiente 1 o 0.

Según la ecuación (5.3.7) y los datos de la tabla 3.2, el lector puede verificar fácilmente que el intervalo de confianza para $\beta_1$ a 95% en este ejemplo es:

$$-1.8871 \leq \beta_1 \leq 1.8583 \tag{5.3.11}$$

Una vez más, se debe tener cuidado al interpretar este intervalo de confianza. En 95 de cada 100 casos, los intervalos como la ecuación (5.3.11) contendrán al verdadero valor de $\beta_1$; la probabilidad de que este intervalo fijo en particular incluya al verdadero $\beta_1$ es de 1 o 0.

### Intervalo de confianza para $\beta_1$ y $\beta_2$ simultáneamente

Hay ocasiones en que se necesita construir un *intervalo de confianza conjunto* para $\beta_1$ y $\beta_2$ tal que, para un coeficiente de confianza $(1 - \alpha)$ de, por ejemplo, 95%, tanto $\beta_1$ como $\beta_2$ caigan al mismo tiempo dentro de ese intervalo. Como este tema es complejo, el lector quizá desee consultar referencias apropiadas.[5] (Estudiaremos brevemente este tema en los capítulos 8 y 10.)

## 5.4 Intervalo de confianza para $\sigma^2$

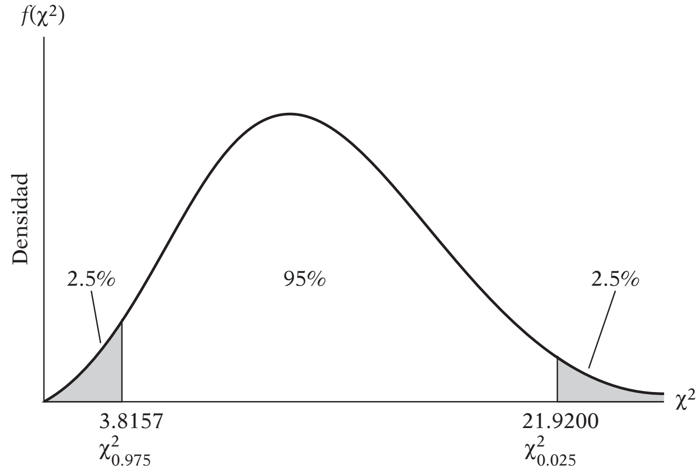
Como señalamos en el capítulo 4, sección 4.3, según el supuesto de normalidad, la variable

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

[4] Debido a errores de redondeo en la tabla 3.2, las respuestas que se proporcionan a continuación pueden no coincidir exactamente con las obtenidas mediante un paquete estadístico.

[5] Hay un análisis accesible en John Neter, William Wasserman y Michael H. Kutner, *Applied Linear Regression Models,* Richard D. Irwin, Homewood, Illinois, 1983, capítulo 5.

**FIGURA 5.1**
Intervalo de confianza a 95% para $\chi^2$ (11 gl).




sigue la distribución $\chi^2$ con $n - 2$ gl.[6] Por tanto, con la distribución $\chi^2$ se establece el intervalo de confianza para $\sigma^2$

$$\Pr\left(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}\right) = 1 - \alpha \tag{5.4.2}$$

donde el valor de $\chi^2$ en medio de esta doble desigualdad es igual a la planteada en (5.4.1), y donde $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ son dos valores de $\chi^2$ (los valores **críticos** $\chi^2$) obtenidos de la tabla ji cuadrada para $n - 2$ gl de manera que ellos cortan $100(\alpha/2)\%$ de las áreas de las colas de la distribución $\chi^2$, como se muestra en la figura 5.1.

Al sustituir $\chi^2$ de (5.4.1) en (5.4.2) y reorganizar los términos, obtenemos

$$\Pr\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha \tag{5.4.3}$$

que da el intervalo de confianza a $100(1 - \alpha)\%$ para $\sigma^2$.

Continuamos con el ejemplo de salarios y nivel de escolaridad: se encontró en la tabla 3.2 que para los datos se tiene $\hat{\sigma}^2 = 0.8936$. Si seleccionamos $\alpha$ de 5%, la tabla ji cuadrada para 11 gl da los siguientes valores críticos: $\chi^2_{0.025} = 21.9200$ y $\chi^2_{0.975} = 3.8157$. Estos valores muestran que la probabilidad de que un valor ji cuadrada sea superior a 21.9200 es 2.5%, y la de 3.8157 es 97.5%. Por consiguiente, el intervalo entre estos dos valores es el intervalo de confianza para $\chi^2$ a 95%, como se aprecia en el diagrama de la figura 5.1. (Observe la característica asimétrica de la distribución ji cuadrada.)

Al sustituir los datos del ejemplo en (5.4.3), el lector verificará que el intervalo de confianza para $\sigma^2$ a 95% es el siguiente:

$$0.4484 \leq \sigma^2 \leq 2.5760 \tag{5.4.4}$$

**La interpretación de este intervalo es la siguiente**: Si establecemos límites de confianza a 95% sobre $\sigma^2$ y afirma *a priori* que entre estos límites caerá el verdadero $\sigma^2$, acertaremos, a la larga, 95% de las veces.

[6] Para una demostración, véase Robert V. Hogg y Allen T. Craig, *Introduction to Mathematical Statistics*, 2a. ed., Macmillan, Nueva York, 1965, p. 144.

## 5.5 Prueba de hipótesis: comentarios generales

Tras estudiar el problema de la estimación puntual y por intervalos, ahora consideraremos el tema de las pruebas de hipótesis. En esta sección analizaremos brevemente algunos aspectos generales de este tema; el **apéndice A** proporciona algunos detalles adicionales.

El problema de las pruebas de hipótesis estadísticas puede plantearse sencillamente de la siguiente manera: *¿es compatible o no lo es una observación o un hallazgo dados, según algunas hipótesis planteadas?* La palabra "compatible" se utiliza aquí en el sentido de que la observación es lo "bastante" cercana al valor hipotético, de forma que no se rechaza la hipótesis planteada. Así, si alguna teoría o experiencia previa lleva a creer que el verdadero coeficiente de la pendiente $\beta_2$ en el ejemplo salarios y nivel de escolaridad es la unidad, ¿es el $\hat{\beta}_2 = 0.724$ obtenido de la muestra de la tabla 3.2 consistente con la hipótesis planteada? De ser así, no se rechaza la hipótesis; de lo contrario, se puede rechazar.

En el lenguaje de estadística, la hipótesis planteada se conoce como **hipótesis nula,** y se denota con el símbolo $H_0$. La hipótesis nula suele probarse frente a una **hipótesis alternativa** (también conocida como **hipótesis mantenida**) denotada con $H_1$, que puede plantear, por ejemplo, que el verdadero $\beta_2$ es diferente a la unidad. La hipótesis alternativa puede ser **simple** o **compuesta**.[7] Por ejemplo, $H_1$:$\beta_2 = 1.5$ es una hipótesis simple, pero $H_1$:$\beta_2 \neq 1.5$ es una hipótesis compuesta.

La teoría de pruebas de hipótesis se refiere al diseño de reglas o procedimientos que permitan decidir si se rechaza o no la hipótesis nula. Hay dos métodos *mutuamente complementarios* para diseñar tales reglas: el **intervalo de confianza** y la **prueba de significancia**. Estos dos enfoques plantean que la variable (el estadístico o estimador) en consideración sigue alguna distribución de probabilidad y que la prueba de hipótesis establece afirmaciones sobre el (los) valor(es) del (los) parámetro(s) de tal distribución. Por ejemplo, sabemos que, con el supuesto de normalidad, $\hat{\beta}_2$ está normalmente distribuida con media igual a $\beta_2$ y varianza dada por (4.3.5). Si formulamos la hipótesis de que $\beta_2 = 1$, se hace una afirmación sobre uno de los parámetros de la distribución normal, por ejemplo, la media. La mayoría de las hipótesis estadísticas que se encuentran en este texto será de este tipo, con afirmaciones sobre uno o más valores de los parámetros de algunas distribuciones de probabilidad supuestas, como la normal, $F$, $t$ o $\chi^2$. En las dos secciones siguientes estudiaremos la forma de lograrlo.

## 5.6 Pruebas de hipótesis: método del intervalo de confianza

### Prueba bilateral o de dos colas

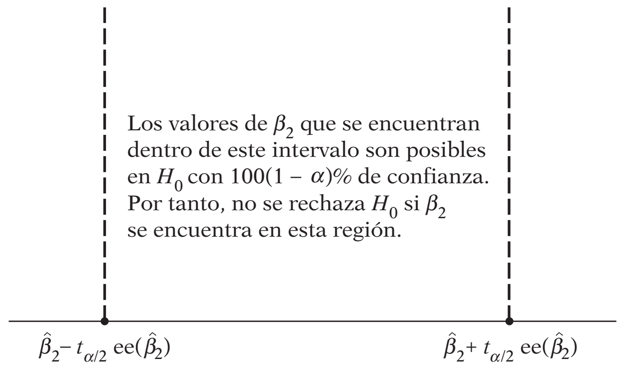
Para ilustrar el enfoque del intervalo de confianza, una vez más nos referiremos al ejemplo de salarios y nivel de escolaridad. Por los resultados de la regresión obtenidos en la ecuación (3.6.1), sabemos que el coeficiente de pendiente es 0.7240. Supongamos que se postula que

$$H_0: \beta_2 = 0.5$$
$$H_1: \beta_2 \neq 0.5$$

es decir, el verdadero coeficiente de la pendiente es 0.5 según la hipótesis nula, pero menor o mayor que 0.5 según la hipótesis alternativa. La hipótesis nula es una hipótesis simple, mientras

[7] Una hipótesis estadística se denomina **hipótesis simple** si especifica el(los) valor(es) preciso(s) del(los) parámetro(s) de una función de densidad de probabilidad (fdp); de lo contrario, se llama **hipótesis compuesta**. Por ejemplo, en la fdp normal $(1/\sigma\sqrt{2\pi})\exp\{-\frac{1}{2}[(X-\mu)/\sigma]^2\}$, si afirmamos que $H_1$:$\mu$ y $\sigma = 2$, se trata de una hipótesis simple; pero si $H_1$:$\mu = 15$ y $\sigma > 15$, es una hipótesis compuesta, porque la desviación estándar no tiene un valor específico.

**FIGURA 5.2**
Intervalo de confianza para $\beta_2$ a $100(1-\alpha)\%$.




que la hipótesis alternativa es compuesta; y, en la práctica, se conoce como **hipótesis bilateral**. Con mucha frecuencia, dicha hipótesis alternativa bilateral refleja el hecho de que no se tiene una expectativa *a priori* o teórica sólida sobre la dirección en la cual debe moverse la hipótesis alternativa respecto de la hipótesis nula.

¿Es el $\hat{\beta}_2$ observado compatible con $H_0$? Para responder, consultemos el intervalo de confianza (5.3.9). Sabemos que, a la larga, los intervalos como (0.5700, 0.8780) contendrán al verdadero $\beta_2$ con una probabilidad de 95%. En consecuencia, a la larga (es decir, en muestreo repetido), tales intervalos proporcionan un recorrido o límites dentro de los cuales puede encontrarse el verdadero $\beta_2$ con un coeficiente de confianza de 95%. Así, el intervalo de confianza proporciona un conjunto de hipótesis nulas posibles. Por consiguiente, si el $\beta_2$ en $H_0$ se encuentra dentro del intervalo de confianza $100(1-\alpha)\%$, no se rechaza la hipótesis nula; si se encuentra por fuera del intervalo, se puede rechazar.[8] Este intervalo se ilustra esquemáticamente en la figura 5.2.

***Regla de decisión*** Construya un intervalo de confianza para $\beta_2$ a $100(1-\alpha)\%$. Si el $\beta_2$ en $H_0$ se encuentra dentro de este intervalo de confianza, no rechace $H_0$, pero si está fuera del intervalo, rechace $H_0$.

Según esta regla, para el ejemplo hipotético, es claro que $H_0{:}\beta_2 = 0.5$ se encuentra fuera del intervalo de confianza a 95% dado en (5.3.9). Por consiguiente, rechazamos la hipótesis de que la verdadera pendiente sea 0.5, con 95% de confianza. Si la hipótesis nula fuera cierta, la probabilidad de obtener por casualidad un valor de pendiente igual a 0.7240 es, como máximo, alrededor de 5%, una probabilidad pequeña.

En estadística, cuando se rechaza la hipótesis nula, se dice que el hallazgo es **estadísticamente significativo**. Por otra parte, cuando no se hace, se dice que el resultado no es **estadísticamente significativo**.

Algunos autores utilizan frases como "muy significativo desde un punto de vista estadístico". Con este término, por lo general se refieren a que, cuando rechazan la hipótesis nula, la probabilidad de cometer un error tipo I (por ejemplo, $\alpha$) es un número pequeño, usualmente 1%. Pero, como demostrará el análisis del **valor *p*** en la sección 5.8, es mejor dejar que el investigador califique el hallazgo estadístico como "significativo", "moderadamente significativo" o "muy significativo".

[8] Siempre tenga en mente que hay una posibilidad de $100\alpha$ de que el intervalo de confianza no contenga a $\beta_2$ en $H_0$ aunque la hipótesis esté correcta. En pocas palabras, hay una posibilidad de $100\alpha$ de cometer un **error tipo I**. Así, si $\alpha = 0.05$, hay una posibilidad de 5% de rechazar la hipótesis nula aunque sea verdadera.

### Prueba unilateral o de una cola

Algunas veces tenemos una expectativa *a priori* o teórica sólida (o existen expectativas basadas en algún trabajo empírico previo) de que la hipótesis alternativa es unilateral o unidireccional, en lugar de ser bilateral o de dos colas, como acabamos de analizar. Así, para el ejemplo de los salarios y el nivel de escolaridad, se puede postular que

$$H_0{:}\beta_2 \leq 0.5 \quad \text{y} \quad H_1{:}\beta_2 > 0.5$$

Quizá la teoría económica o el trabajo empírico previo indiquen que la pendiente es mayor que 0.5. Aunque el procedimiento para probar esta hipótesis se deriva fácilmente de (5.3.5), se explica mejor el mecanismo real en términos del método de prueba de significancia siguiente.[9]

## 5.7 Pruebas de hipótesis: enfoque de la prueba de significancia

### Prueba de significancia de los coeficientes de regresión: la prueba *t*

Un *enfoque alterno pero complementario* al de intervalos de confianza para probar hipótesis estadísticas es el **método de la prueba de significancia**, desarrollado en forma independiente por R. A. Fisher y conjuntamente por Neyman y Pearson.[10] **En términos generales, una prueba de significancia es un procedimiento que utiliza los resultados muestrales para verificar la verdad o falsedad de una hipótesis nula.** La idea básica de las pruebas de significancia es la de un **estadístico de prueba** (un estimador) y su distribución muestral según la hipótesis nula. La decisión de aceptar o rechazar $H_0$ se toma con base en el valor del estadístico de prueba obtenido con los datos disponibles.

Como ilustración, recuerde que, según el supuesto de normalidad, la variable

$$\begin{aligned} t &= \frac{\hat{\beta}_2 - \beta_2}{\text{ee}\,(\hat{\beta}_2)} \\ &= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \end{aligned} \tag{5.3.2}$$

sigue la distribución $t$ con $n-2$ gl. Si el valor del verdadero $\beta_2$ se especifica con la hipótesis nula, el valor $t$ de (5.3.2) se calcula fácilmente a partir de la muestra disponible y, por consiguiente, sirve como estadístico de prueba. Y como este estadístico de prueba sigue una distribución $t$, caben afirmaciones sobre los intervalos de confianza como la siguiente:

$$\Pr\left[-t_{\alpha/2} \leq \frac{\hat{\beta}_2 - \beta_2^*}{\text{ee}\,(\hat{\beta}_2)} \leq t_{\alpha/2}\right] = 1 - \alpha \tag{5.7.1}$$

donde $\beta_2^*$ es el valor de $\beta_2$ en $H_0$ y $-t_{\alpha/2}$ y $t_{\alpha/2}$ son los valores de $t$ (los valores **críticos** de $t$) obtenidos de la tabla $t$ para un nivel de significancia $(\alpha/2)$ y $n-2$ gl [cf. (5.3.4)]. La tabla $t$ está en el **apéndice D**.

[9] Si desea utilizar el método de intervalos de confianza, elabore un intervalo de confianza *unilateral* o *de una cola* para $\beta_2$ a $(100 - \alpha)$%. ¿Por qué?

[10] Hay detalles en E. L. Lehman, *Testing Statistical Hypotheses*, John Wiley & Sons, Nueva York, 1959.

Reorganizamos (5.7.1) para obtener

$$\Pr\left[\beta_2^* - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \le \hat{\beta}_2 \le \beta_2^* + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2)\right] = 1 - \alpha \qquad (5.7.2)$$

que da el intervalo en el cual se encontrará $\hat{\beta}_2$ con probabilidad $1 - \alpha$, dado $\beta_2 = \beta_2^*$. En el lenguaje de pruebas de hipótesis, el intervalo de confianza a $100(1 - \alpha)\%$ establecido en (5.7.2) se conoce como la **región de aceptación** (de la hipótesis nula), y *la(s) región(es)* que queda(n) fuera del intervalo de confianza se llama(n) **región(es) de rechazo** (de $H_0$) o **región(es) crítica(s)**. Como ya mencionamos, los límites de confianza dados por los puntos extremos del intervalo de confianza se llaman también **valores críticos**.

Ahora se aprecia la estrecha conexión entre los enfoques de intervalo de confianza y prueba de significancia para realizar pruebas de hipótesis al comparar (5.3.5) con (5.7.2). En el procedimiento de intervalo de confianza se trata de establecer un rango o intervalo que tenga una probabilidad determinada de contener al verdadero aunque desconocido $\beta_2$, mientras que en el enfoque de prueba de significancia se somete a hipótesis algún valor de $\beta_2$ y se ve si el $\hat{\beta}_2$ calculado se encuentra dentro de límites (de confianza) razonables alrededor del valor sometido a hipótesis.

Veamos de nuevo el ejemplo de los salarios y el nivel de escolaridad. Sabemos que $\hat{\beta}_2 = 0.7240$, ee $(\hat{\beta}_2) = 0.0700$ y gl = 11. Si suponemos $\alpha = 5\%$, $t_{\alpha/2} = 2.201$.

Si suponemos que $H_0$: $\beta_2 = \beta_2^* = 0.5$ y $H_1$: $\beta_2 \ne 0.5$, (5.7.2) se convierte en

$$\Pr\,(0.3460 \le \hat{\beta}_2 \le 0.6540) \qquad (5.7.3)^{11}$$

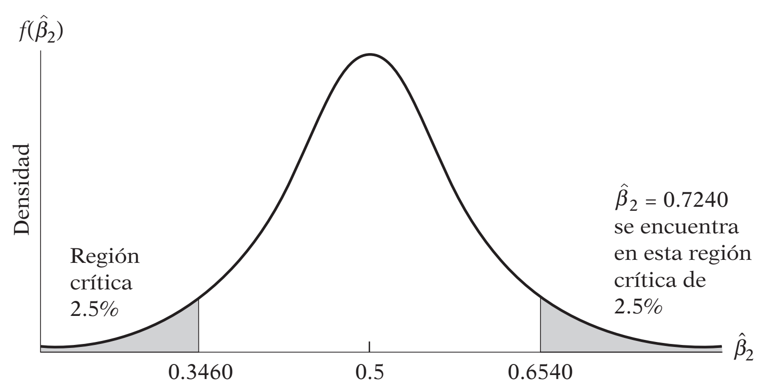
como se muestra en el diagrama de la figura 5.3.

En la práctica, no hay necesidad de estimar (5.7.2) explícitamente. Se calcula el valor de $t$ del centro de la doble desigualdad dada en (5.7.1) y se ve si cae entre los valores críticos $t$ o fuera de ellos. Para el ejemplo,

$$t = \frac{0.7240 - 0.5}{0.0700} = 3.2 \qquad (5.7.4)$$

valor que con claridad se encuentra en la región crítica de la figura 5.4. La conclusión se mantiene; es decir, rechazamos $H_0$.

**FIGURA 5.3** Intervalo de confianza a 95% para $\hat{\beta}_2$ según la hipótesis de que $\beta_2 = 0.5$.




[11] En la sección 5.2, punto 4, se afirmó que *no se puede* decir que la probabilidad de que el intervalo fijo (0.4268, 0.5914) incluya el verdadero $\beta_2$, sea de 95%. Pero sí se puede hacer la afirmación probabilística dada en (5.7.3) porque $\hat{\beta}_2$, al ser un estimador, es una variable aleatoria.

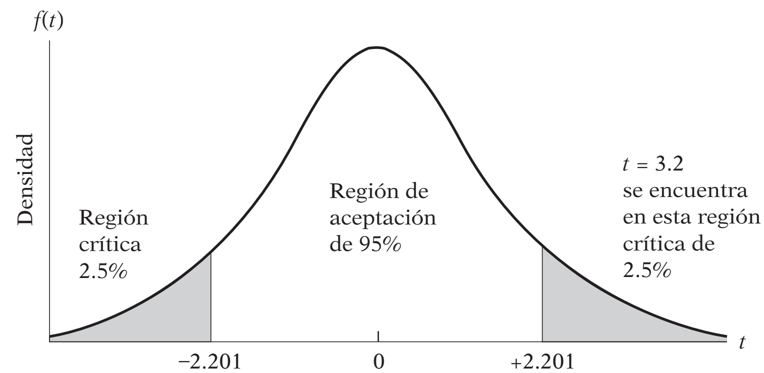



**FIGURA 5.4**
Intervalo de confianza a 95% para $t$(11 gl).

Observe que si el $\beta_2$ ($= \hat{\beta}_2$) estimado es igual al $\beta_2$ hipotético, el valor $t$ en (5.7.4) será cero. Sin embargo, en la medida en que el valor de $\beta_2$ estimado se aleje del valor hipotético de $\beta_2$, el $|t|$ (es decir, el valor absoluto de *t; nota*: $t$ puede ser positivo o negativo) será cada vez mayor. *Por consiguiente, un valor "grande" de* $|t|$ *será evidencia en contra de la hipótesis nula*. Siempre se puede utilizar la tabla $t$ para determinar si un valor $t$ particular es grande o pequeño; la respuesta, como sabemos, depende de los grados de libertad igual que de la probabilidad del error tipo I que estemos dispuestos a aceptar. Como se observa en la tabla $t$ del **apéndice D** (tabla D.2), para cualquier valor dado de gl, la probabilidad de obtener un valor de $|t|$ cada vez mayor es progresivamente menor. Por tanto, para 20 gl, la probabilidad de obtener un valor $|t|$ mayor o igual a 1.725 es 0.10 o 10%, pero para los mismos gl, la probabilidad de obtener un valor $|t|$ mayor o igual a 3.552 es tan sólo 0.002, o 0.2%.

Como utilizamos la distribución $t$, el anterior procedimiento de prueba se llama, apropiadamente, **prueba *t*. En el lenguaje de las pruebas de significancia, se dice que un estadístico es estadísticamente significativo si el valor del estadístico de prueba cae en la región crítica. En este caso, se rechaza la hipótesis nula. De la misma manera, se dice que una prueba no es estadísticamente significativa si el valor del estadístico de prueba cae en la región de aceptación.** En esta situación, no se rechaza la hipótesis nula. En el ejemplo, la prueba $t$ es significativa y por tanto se rechaza la hipótesis nula.

Antes de concluir la exposición de pruebas de hipótesis, observe que este procedimiento de prueba se conoce como procedimiento de las pruebas de significancia **bilaterales,** o **de dos colas**, pues se consideran las dos colas extremas de la distribución de probabilidad relevante, las regiones de rechazo, y se rechaza la hipótesis nula si cae en cualquiera de ellas. Esto sucede porque la $H_1$ era una hipótesis compuesta de dos lados; $\beta_2 \neq 0.5$ significa que $\beta_2$ es mayor o menor que 0.5. Sin embargo, supongamos que la experiencia indica que la pendiente es mayor que 0.5. En este caso tenemos: $H_0$: $\beta_2 \leq 0.5$ y $H_1$:$\beta_2 > 0.5$. Aunque $H_1$ es aún una hipótesis compuesta, ahora es unilateral. Para probar esta hipótesis se utiliza una **prueba de una cola** (la cola derecha), como se ve en la figura 5.5. (Véase también el análisis en la sección 5.6.)

El procedimiento de prueba es similar al anterior excepto que el límite de confianza superior o valor crítico corresponde ahora a $t_\alpha = t_{0.05}$, es decir, al nivel de 5%. Como indica la figura 5.5, en este caso no es preciso considerar la cola inferior de la distribución $t$. La utilización de una prueba de significancia de una o dos colas dependerá de la forma como se formule la hipótesis alternativa, la cual, a su vez, puede depender de algunas consideraciones *a priori* o de experiencia empírica previa. (Otras consideraciones se dan en la sección 5.8.)

En la tabla 5.1 se presenta un resumen del método de la prueba $t$ de significancia para pruebas de hipótesis.

**FIGURA 5.5**
Prueba de significancia de una cola.

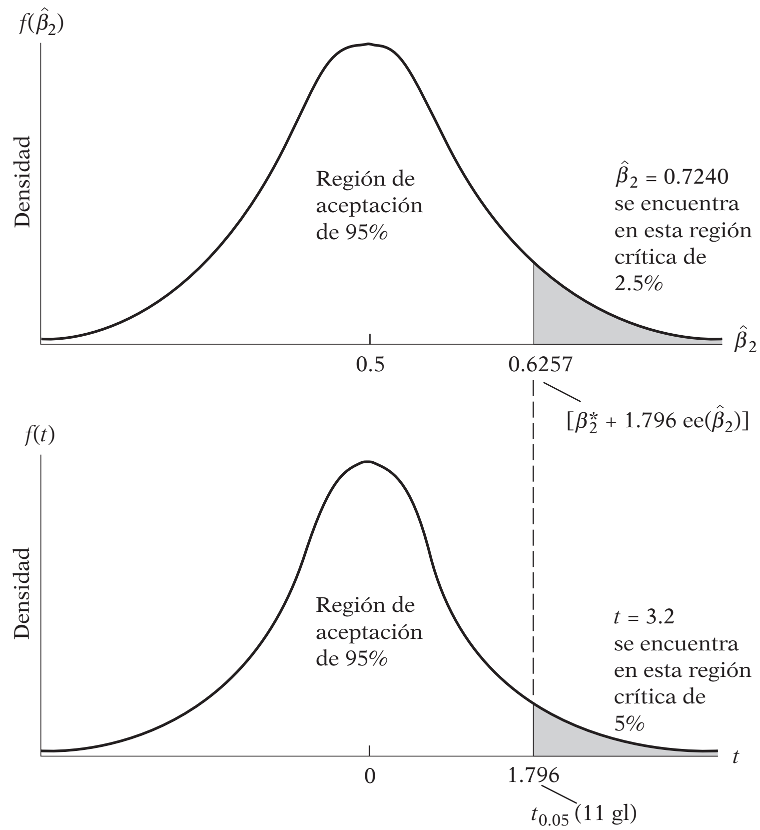



**TABLA 5.1**
**La prueba *t* de significancia: reglas de decisión**

| Tipo de hipótesis | $H_0$: hipótesis nula | $H_1$: hipótesis alternativa | Regla de decisión: rechazar $H_0$ si |
|---|---|---|---|
| Dos colas | $\beta_2 = \beta_2^*$ | $\beta_2 \neq \beta_2^*$ | $\lvert t \rvert > t_{\alpha/2,\text{gl}}$ |
| Cola derecha | $\beta_2 \leq \beta_2^*$ | $\beta_2 > \beta_2^*$ | $t > t_{\alpha,\text{gl}}$ |
| Cola izquierda | $\beta_2 \geq \beta_2^*$ | $\beta_2 < \beta_2^*$ | $t < -t_{\alpha,\text{gl}}$ |

*Notas:* $\beta_2^*$ es el valor numérico hipotético de $\beta_2$.
$|t|$ significa el valor absoluto de $t$.
$t_\alpha$ o $t_{\alpha/2}$ significa el valor crítico de $t$ al nivel de significancia $\alpha$ o $\alpha/2$.
gl: grados de libertad, $(n-2)$ para el modelo con dos variables, $(n-3)$ para el modelo con tres variables, y así sucesivamente.
Para probar hipótesis sobre $\beta_1$ se sigue un procedimiento similar.

## Prueba de significancia de $\sigma^2$: la prueba $\chi^2$

Como otro ejemplo de la metodología de las pruebas de significancia, considere la siguiente variable:

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

la cual, como ya mencionamos, sigue una distribución $\chi^2$ con $n-2$ gl. Para el ejemplo hipotético, $\hat{\sigma}^2 = 0.8937$ y gl $= 11$. Si se postula que $H_0$: $\sigma^2 = 0.6$ frente a $H_1$:$\sigma^2 \neq 0.6$, la ecuación (5.4.1) proporciona el estadístico de prueba para $H_0$. Al sustituir los valores apropiados en (5.4.1), se descubre que, con $H_0$, $\chi^2 = 16.3845$. Si suponemos que $\alpha = 5\%$, los valores críticos $\chi^2$

**TABLA 5.2**
**Resumen de la prueba $\chi^2$**

| $H_0$: hipótesis nula | $H_1$: hipótesis alternativa | Región crítica: rechazar $H_0$ si |
|---|---|---|
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 > \sigma_0^2$ | $\frac{\text{gl}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{gl}}$ |
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 < \sigma_0^2$ | $\frac{\text{gl}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{gl}}$ |
| $\sigma^2 = \sigma_0^2$ | $\sigma^2 \neq \sigma_0^2$ | $\frac{\text{gl}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{gl}}$ o $< \chi^2_{(1-\alpha/2),\text{gl}}$ |

*Nota:* $\sigma_0^2$ es el valor de $\sigma^2$ según la hipótesis nula. El primer subíndice asociado a $\chi^2$ en la última columna es el nivel de significancia, en tanto que el segundo indica los grados de libertad. Éstos son los valores críticos ji cuadrada. Observe que, si el modelo de regresión es de dos variables, los gl son $(n - 2)$, si el modelo de regresión es de tres variables son $(n - 3)$ y así sucesivamente.

son 3.81575 y 21.9200. Como el valor $\chi^2$ calculado cae dentro de estos límites, los datos apoyan la hipótesis nula y no la rechazamos. (Véase la figura 5.1.) Este procedimiento de prueba se denomina **prueba de significancia ji cuadrada**. El método de la prueba de significancia $\chi^2$ para la prueba de hipótesis se resume en la tabla 5.2.

## 5.8 Prueba de hipótesis: algunos aspectos prácticos

### Significado de "aceptar" o "rechazar" una hipótesis

Si, con base en una prueba de significancia, por ejemplo, la prueba *t*, decidimos "aceptar" la hipótesis nula, todo lo que se afirma es que, con base en la evidencia dada por la muestra, no existe razón para rechazarla; no se sostiene que la hipótesis nula sea verdadera con absoluta certeza. ¿Por qué? Para responder esto, regresemos al ejemplo de los salarios y los niveles de escolaridad y supongamos que $H_0$: $\beta_2 = 0.70$. Ahora, el valor estimado de la pendiente es $\hat{\beta}_2 = 0.7241$ con un ee $(\hat{\beta}_2) = 0.0701$. En seguida, con base en la prueba *t*, se obtiene que $t = \frac{(0.7241 - 0.7)}{0.0701} = 0.3438$, que no es significativo, por ejemplo, en $\alpha = 5\%$. Por consiguiente, se dice que "aceptamos" $H_0$. Pero ahora supongamos que $H_0$: $\beta_2 = 0.6$. Al aplicar de nuevo la prueba *t*, se obtiene $t = \frac{(0.7241 - 0.6)}{0.0701} = 1.7703$, que tampoco es estadísticamente significativo. Entonces, ahora se dice que "aceptamos" esta $H_0$. ¿Cuál de estas dos hipótesis nulas es la "verdadera"? No sabemos. Por consiguiente, al "aceptar" una hipótesis nula siempre se debe tener presente que puede existir otra hipótesis nula igualmente compatible con los datos. Es preferible, por tanto, decir que se *puede* aceptar la hipótesis nula en lugar de decir que la aceptamos. Mejor aún,

> …de la misma manera que en un tribunal se pronuncia un veredicto de "no culpable" en lugar de decir "inocente", así la conclusión de una prueba estadística es la de "no rechazar" en lugar de "aceptar".[12]

[12] Jan Kmenta, *Elements of Econometrics*, Macmillan, Nueva York, 1971, p. 114.

## Hipótesis nula "cero" y regla práctica "2*t*"

La hipótesis nula que es objeto frecuente de prueba en el trabajo empírico es $H_0$: $\beta_2 = 0$, es decir, el coeficiente de la pendiente es cero. Esta hipótesis nula de "cero" es un mecanismo para establecer si $Y$ tiene relación con $X$, la variable explicativa. Si, para empezar, no existe relación entre $Y$ y $X$, entonces no tiene caso probar una hipótesis como $\beta_2 = 0.3$ o cualquier otro valor.

Esta hipótesis nula se prueba fácilmente con los enfoques de intervalos de confianza o prueba $t$ estudiados en las secciones anteriores. Pero, con mucha frecuencia, tales pruebas formales se abrevian con la regla de significancia "2*t*", que puede expresarse así:

***Regla práctica "2t"***

Si el número de grados de libertad es 20 o más, y si $\alpha$, el nivel de significancia, se fija en 0.05, se rechaza la hipótesis nula $\beta_2 = 0$ si el valor de $t$ [ $= \hat{\beta}_2/\text{ee}\,(\hat{\beta}_2)$] calculado a partir de (5.3.2) es superior a 2 en valor absoluto.

El razonamiento de esta regla no es muy difícil. De (5.7.1) se sabe que rechazaremos $H_0$: $\beta_2 = 0$ si

$$t = \hat{\beta}_2/\text{ee}\,(\hat{\beta}_2) > t_{\alpha/2} \qquad \text{cuando } \hat{\beta}_2 > 0$$

o

$$t = \hat{\beta}_2/\text{ee}\,(\hat{\beta}_2) < -t_{\alpha/2} \qquad \text{cuando } \hat{\beta}_2 < 0$$

o cuando

$$|t| = \left| \frac{\hat{\beta}_2}{\text{ee}\,(\hat{\beta}_2)} \right| > t_{\alpha)2} \tag{5.8.1}$$

para los grados de libertad apropiados.

Ahora, si examinamos la tabla $t$ del **apéndice D**, se ve que, para gl alrededor de 20 o más, un valor calculado $t$ mayor que 2 (en términos absolutos), por ejemplo, 2.1, es estadísticamente significativo al nivel de 5%, lo cual implica rechazo de la hipótesis nula. Por consiguiente, si se descubre que para 20 o más gl el valor $t$ calculado es 2.5 o 3, ni siquiera hay que consultar la tabla $t$ para asegurar la significancia del coeficiente de la pendiente estimada. Por supuesto, siempre puede consultar la tabla $t$ para obtener el nivel preciso de significancia. Sin embargo, esto debe hacerse siempre que los gl sean inferiores a, por ejemplo, 20.

A propósito, observe que si se está probando la hipótesis unilateral $\beta_2 = 0$ respecto de $\beta_2 > 0$ o $\beta_2 < 0$, se debe rechazar la hipótesis nula si

$$|t| = \left| \frac{\hat{\beta}_2}{\text{ee}\,(\hat{\beta}_2)} \right| > t_{\alpha} \tag{5.8.2}$$

Si fijamos $\alpha$ en 0.05, en la tabla $t$ se observa que, para 20 o más gl, un valor $t$ mayor que 1.73 es estadísticamente significativo al nivel de significancia de 5% (de una cola). Por tanto, siempre que un valor $t$ exceda, por ejemplo, de 1.8 (en términos absolutos) y los gl sean 20 o más, no es necesario consultar la tabla $t$ para la significancia estadística del coeficiente observado. Es claro que, si se escoge $\alpha$ igual a 0.01 o cualquier otro nivel, habrá que decidir sobre el valor apropiado de $t$ como valor crítico de referencia. Pero a estas alturas, el lector debe ser capaz de hacerlo.

## Formación de las hipótesis nula y alternativa[13]

Con las hipótesis nula y alternativa, probar su significancia estadística no debe seguir siendo un misterio. Pero, ¿cómo se formulan estas hipótesis? No existen reglas específicas. Muy a menudo, el fenómeno en estudio sugerirá la forma de las hipótesis nula y alternativa. Por ejemplo, se pide estimar la línea del mercado de capitales (LMC) de la teoría de portafolios, que postula que $E_i = \beta_1 + \beta_2\sigma_i$, donde $E$ = rendimiento esperado del portafolio y $\sigma$ = la desviación estándar del rendimiento, una medida de riesgo. Como se espera que la ganancia y el riesgo estén relacionados positivamente, entre mayor sea el riesgo, más alta será la ganancia; la hipótesis alternativa natural a la hipótesis nula, $\beta_2 = 0$, sería $\beta_2 > 0$. Es decir, no se considerarán valores de $\beta_2$ menores de cero.

Pero considere el caso de la demanda de dinero. Como demostraremos más adelante, un determinante importante de la demanda de dinero es el ingreso. Estudios anteriores de las funciones de demanda de dinero revelan que la elasticidad ingreso de la demanda de dinero (el cambio porcentual en la demanda de dinero por un cambio porcentual de 1% en el ingreso) suele ubicarse en un rango de 0.7 a 1.3. Por consiguiente, en un nuevo estudio de la demanda de dinero, si se postula que el coeficiente $\beta_2$ de la elasticidad ingreso es 1, la hipótesis alternativa puede ser que $\beta_2 \neq 1$, una hipótesis alternativa bilateral.

Así, las expectativas teóricas o el trabajo empírico previo o ambos pueden ser la base para formular hipótesis. Sin embargo, sin importar la forma de postular hipótesis, *es en extremo importante que el investigador plantee estas hipótesis antes de la investigación empírica*. De lo contrario, él o ella serán culpables de razonamientos circulares o profecías autocumplidas. Es decir, si se formula la hipótesis después de examinar los resultados empíricos, puede presentarse la tentación de formular la hipótesis de manera que justifique los resultados. Deben evitarse estas prácticas a cualquier costo, al menos para salvar la objetividad científica. Recuerde la cita de Stigler al principio de este capítulo.

## Selección del nivel de significancia $\alpha$

Del análisis expuesto hasta ahora, debe tenerse claro que rechazar o no una hipótesis nula depende de $\alpha$, el nivel de significancia o *probabilidad de cometer un error **tipo I***, o sea, la probabilidad de rechazar la hipótesis cuando es verdadera. En el **apéndice A** se analiza en detalle la naturaleza del llamado error tipo I, su relación con el *error **tipo II*** (la probabilidad de aceptar la hipótesis cuando es falsa) y la razón por la cual la estadística clásica suele centrarse en el error tipo I. Pero incluso entonces, ¿por qué $\alpha$ se fija generalmente en los niveles de 1, 5 o, cuanto mucho, 10%? De hecho, no hay nada sagrado acerca de estos valores; cualquier otro valor sería por igual apropiado.

En un libro introductorio como éste no es posible analizar a fondo la razón por la cual se escogen los niveles de significancia 1, 5 o 10%, pues nos llevaría al campo de la toma de decisiones estadísticas, de suyo una disciplina completa. Sin embargo, sí podemos ofrecer un breve resumen. Como estudiamos en el **apéndice A**, para un tamaño de muestra dada, si tratamos de reducir un *error tipo I*, aumenta un *error tipo II,* y viceversa. Es decir, con el tamaño de la muestra, si tratamos de reducir la probabilidad de rechazar la hipótesis cuando es verdadera, se aumenta al mismo tiempo la probabilidad de aceptarla cuando es falsa. Por tanto, dado el tamaño de la mues-

[13] Hay una exposición interesante sobre la formulación de hipótesis en J. Bradford De Long y Kevin Lang, "Are All Economic Hypotheses False?", *Journal of Political Economy*, vol. 100, núm. 6, 1992, pp. 1257-1272.

tra, existe una compensación entre estos dos tipos de error. Ahora, la única forma de decidir sobre esta compensación es encontrar los costos relativos de ambos tipos de error. Entonces,

> Si el error de rechazar la hipótesis nula cuando es verdadera (error tipo I) es costoso en comparación con el error de no rechazar la hipótesis nula cuando es falsa (error tipo II), será razonable fijar la probabilidad de ocurrencia del primer tipo de error en niveles bajos. Si, por otra parte, el costo de incurrir en el error tipo I es bajo comparado con el costo de cometer el error tipo II, se justificará que la probabilidad del primer tipo de error sea alta (lo que reduce la posibilidad de incurrir en el segundo tipo de error).[14]

Desde luego, el problema es que pocas veces se conocen los costos de cometer los dos tipos de error. Por tanto, los econometristas tienen por costumbre fijar el valor de $\alpha$ en niveles de 1, 5 o 10% como máximo, y escogen un estadístico de prueba que haga que la probabilidad de cometer un error tipo II sea lo más pequeña posible. Como uno menos la probabilidad de cometer un error tipo II se conoce como la **potencia de la prueba**, este procedimiento equivale a maximizar la potencia de la prueba. (Véase en el **apéndice A** un análisis de la potencia de una prueba.)

Afortunadamente, el problema relacionado con la selección del valor apropiado de $\alpha$ se evita al emplear lo que se conoce como ***valor p*** del estadístico de prueba, que analizamos a continuación.

## Nivel exacto de significancia: Valor *p*

Como recién mencionamos, el talón de Aquiles del método clásico de la prueba de hipótesis es su arbitrariedad al seleccionar $\alpha$. Una vez obtenido un estadístico de prueba (es decir, el estadístico $t$) en un ejemplo dado, ¿por qué no tan sólo consultar la tabla estadística adecuada y encontrar la probabilidad real de obtener un valor del estadístico de prueba tan grande o mayor que el obtenido en el ejemplo? Esta probabilidad se denomina **valor *p*** (es decir, **valor de probabilidad**), también conocido como **nivel observado** o **exacto de significancia,** o **probabilidad exacta de cometer un error tipo I**. Más técnicamente, el valor $p$ se define como **nivel de significancia más bajo al cual puede rechazarse una hipótesis nula**.

Para ilustrar, retomemos el ejemplo de los salarios y el nivel de escolaridad. Con la hipótesis nula de que el verdadero coeficiente del nivel de escolaridad es 0.5, se obtuvo un valor $t$ de 3.2 en la ecuación (5.7.4). ¿Cuál es el valor $p$ de obtener un valor $t$ igual o superior a 3.2? En la tabla $t$ del **apéndice D** se observa que, para 11 gl, la probabilidad de obtener tal valor $t$ debe estar muy por debajo de 0.005 (una cola) o 0.010 (dos colas).

Si se usan los paquetes estadísticos Stata o EViews, se calcula que el valor $p$ de obtener un valor $t$ igual o mayor que 3.2 es de 0.00001, es decir, muy pequeño. Éste es el valor $p$ del estadístico observado $t$. Este nivel exacto de significancia del estadístico $t$ es mucho menor que el nivel de significancia que se fija de manera convencional y arbitraria, como 1, 5 o 10%. En realidad, si se utilizara el valor $p$ recién calculado y se rechazara la hipótesis nula de que el verdadero coeficiente de escolaridad es 0.5, la probabilidad de cometer un error tipo I sería más o menos de sólo 1 en 100 000.

Como ya mencionamos, si los datos no apoyan la hipótesis nula, el $|t|$ obtenido con tal hipótesis nula será "grande" y, por consiguiente, el valor $p$ de obtener tal $|t|$ será "pequeño". En otras palabras, para un tamaño de muestra dado, a medida que aumenta $|t|$, el valor $p$ se reduce y, por consiguiente, se rechaza la hipótesis nula con mayor confianza.

¿Cuál es la relación entre el valor $p$ y el nivel de significancia $\alpha$? Si se adquiere el hábito de fijar $\alpha$ igual al valor $p$ de un estadístico de prueba (es decir, el estadístico $t$), entonces no hay conflicto entre estos dos valores. En otros términos, **es mejor dejar de fijar $\alpha$ de forma arbitraria**

[14] Jan Kmenta, *Elements of Econometrics*, Macmillan, Nueva York, 1971, pp. 126-127.

**en algún nivel y tan sólo seleccionar el valor *p* del estadístico de prueba**. Es preferible dejar que el lector decida si debe rechazar la hipótesis nula con el valor *p* dado. Si, en una aplicación, el valor *p* de un estadístico de prueba resulta ser, por ejemplo, 0.145 o 14.5%, y si el lector desea rechazar la hipótesis nula en este nivel (exacto) de significancia, que así sea. No está mal correr el riesgo de equivocarse 14.5% de las veces si se rechaza la hipótesis nula verdadera. De manera similar, como en el ejemplo de los salarios y el nivel de escolaridad, no está mal si el investigador desea escoger un valor *p* cercano a 0.02% y no correr el riesgo de equivocarse más de 2 veces de cada 10 000. Después de todo, algunos investigadores pueden amar el riesgo y otros no.

*En el resto de este texto por lo general citaremos el valor p de un estadístico de prueba dado.* Tal vez algunos lectores deseen fijar $\alpha$ en algún nivel y rechazar la hipótesis nula si el valor *p* es menor que $\alpha$. Es su decisión.

## Significancia estadística y significancia práctica

Recordemos el ejemplo 3.1 y los resultados de la regresión en la ecuación (3.7.1). Esta regresión relaciona el consumo personal (GCP) con el producto interno bruto (PIB) en Estados Unidos de 1960 a 2005; ambas variables se miden en miles de millones de dólares de 2000.

En esta regresión observamos que la propensión marginal a consumir (PMC), es decir, el consumo adicional que produce un dólar adicional de ingreso (medido por el PIB) es de alrededor de 0.72, o 72 centavos. Con los datos de la ecuación (3.7.1), el lector puede verificar que el intervalo de confianza a 95% para la PMC sea (0.7129, 0.7306). (*Nota:* Como hay 44 gl en este problema, no contamos con un valor crítico *t* preciso para estos gl. En consecuencia, el intervalo de confianza a 95% se calcula con la regla práctica *2t*.)

Suponga que alguien afirma que la verdadera PMC es 0.74. ¿Esta cifra difiere de 0.72? Sí, si nos apegamos estrictamente al intervalo de confianza establecido antes.

Pero, ¿cuál es el significado práctico o real del hallazgo? Es decir, ¿qué diferencia hay entre asignar a la PMC un valor de 0.74 o uno de 0.72? ¿Es la diferencia de 0.02 entre las dos PMC muy importante en la práctica?

La respuesta depende de lo que en realidad se haga con estas estimaciones. Por ejemplo, de la macroeconomía se sabe que el multiplicador del ingreso es $1/(1 - \text{PMC})$. Por tanto, si la PMC es 0.72, el multiplicador es 3.57, pero será 3.84 si la PMC es igual a 0.74. Es decir, si el gobierno incrementara su gasto en 1 dólar para sacar la economía de una recesión, el ingreso aumentaría 3.57 dólares, si la PMC fuese 0.72, pero lo hará 3.84 dólares si la PMC es 0.74. Y esa diferencia puede ser crucial para reactivar la economía.

*El punto de toda esta exposición es que no se debe confundir la significancia estadística con la significancia práctica o económica*. Como afirma Goldberger:

> Cuando se especifica una hipótesis nula, digamos $\beta_j = 1$, lo que se busca es que $\beta_j$ esté *cercano* a 1, tan cerca que para todos los propósitos prácticos pueda tratarse *como si fuera* 1. Pero que 1.1 sea "prácticamente lo mismo que" 1.0 es un asunto de economía, no de estadística. El asunto no se resuelve con una prueba de hipótesis, porque el estadístico de prueba $[t = ](b_j - 1)/\hat{\sigma}_{bj}$ mide el coeficiente estimado en unidades de errores estándar, las cuales no tienen significado para medir el parámetro económico $\beta_j - 1$. Puede ser una buena idea reservar el término "significancia" para el concepto estadístico, y adoptar la palabra "sustancial" para el económico.[15]

[15] Arthur S. Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991, p. 240. Observe que $b_j$ es el estimador de MCO de $\beta_j$, y $\hat{\sigma}_{bj}$ es su error estándar. Este enfoque se corrobora en D. N. McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests", *American Economic Review,* vol. 75, 1985, pp. 201-205. Véase también D. N. McCloskey y S. T. Ziliak, "The Standard Error of Regression," *Journal of Economic Literature*, vol. 37, 1996, pp. 97-114.

El punto expresado por Goldberger es importante. A medida que el tamaño de la muestra se hace muy grande, la importancia de los temas relacionados con significancia estadística se reduce mucho, pero los temas de significancia económica adquieren importancia crítica. De hecho, como con muestras grandes se rechazan casi todas las hipótesis nulas, puede haber estudios en los cuales lo único importante sea la magnitud de los valores estimados puntuales.

### Elección entre los enfoques de intervalos de confianza y pruebas de significancia en las pruebas de hipótesis

En la mayor parte de los análisis económicos aplicados, la hipótesis nula postulada hace las veces de comodín, y el objetivo del trabajo empírico es tumbarlo, es decir, rechazar la hipótesis nula. Por tanto, en el ejemplo consumo-ingreso, la hipótesis nula de que la PMC $\beta_2 = 0$ es a todas luces absurda, pero con frecuencia sirve para ejemplificar los resultados empíricos. Parece que a los editores de publicaciones especializadas de renombre no les emociona publicar un trabajo empírico que no rechace la hipótesis nula. De alguna manera, como noticia, es más novedoso el hallazgo de que la PMC sea estadísticamente diferente de cero que el hallazgo de que sea igual a, digamos, 0.7.

Así, J. Bradford De Long y Kevin Lang sostienen que es mejor para los economistas

> …concentrarse en las magnitudes de los coeficientes e informar sobre los niveles de confianza y no sobre las pruebas de significancia. Si todas, o casi todas, las hipótesis nulas son falsas, no es muy sensato concentrarse en averiguar si una estimación es o no distinguible de su valor predicho con la hipótesis nula. En lugar de esto, deseamos saber qué modelos son buenas aproximaciones, para lo cual es necesario conocer los intervalos de los valores de los parámetros excluidos por las estimaciones empíricas.[16]

En resumen, estos autores prefieren el enfoque de intervalos de confianza al de pruebas de significancia. Este consejo puede ser de utilidad para el lector.[17]

## 5.9 Análisis de regresión y análisis de varianza

En esta sección estudiamos el análisis de regresión desde el punto de vista del análisis de varianza, y nos introduciremos en una forma complementaria de mirar el problema de la inferencia estadística.

En el capítulo 3, sección 3.5, elaboramos la siguiente identidad:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \qquad \textbf{(3.5.2)}$$

es decir, SCT = SCE + SCR, la cual fragmenta la suma de cuadrados total (SCT) en dos componentes: la suma de cuadrados explicada (SCE) y la suma de cuadrados de residuos (SCR). El estudio de estos componentes de SCT se conoce como **análisis de varianza** (ANOVA) desde el punto de vista de la regresión.

Asociados con toda suma de cuadrados están sus gl, es decir, el número de observaciones independientes en las que se basa. La SCT tiene $n - 1$ gl porque se pierde 1 gl en el cálculo de la media muestral . La SCR tiene $n - 2$ gl. (¿Por qué?) (*Nota:* Esto sólo es válido para el modelo de regresión con dos variables con presencia del intercepto $\beta_1$.) SCE tiene 1 gl (de nuevo, esto sólo vale para el caso de dos variables), lo cual se deduce de que SCE $= \hat{\beta}_2^2 \sum x_i^2$ sea una función sólo de $\hat{\beta}_2$, pues se conoce $\sum x_i^2$.

[16] Véase su artículo citado en la nota 13, p. 1271.

[17] Para una perspectiva un tanto distinta, véase Carter Hill, William Griffiths y George Judge, *Undergraduate Econometrics*, Wiley & Sons, Nueva York, 2001, p. 108.

**TABLA 5.3**
**Tabla ANOVA para el modelo de regresión con dos variables**

| Fuente de variación | SC* | gl | SCP† |
|---|---|---|---|
| Debido a la regresión (SCE) | $\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$ | 1 | $\hat{\beta}_2^2 \sum x_i^2$ |
| Debido a los residuos (SCR) | $\sum \hat{u}_i^2$ | $n-2$ | $\dfrac{\sum \hat{u}_i^2}{n-2} = \hat{\sigma}^2$ |
| SCT | $\sum y_i^2$ | $n-1$ | |

* SC significa suma de cuadrados.
† Significa suma de cuadrados promedio, la cual se obtiene al dividir SC entre el número de gl.

Reorganicemos las sumas de cuadrados y sus gl asociados en la tabla 5.3, que es la forma estándar de la tabla AOV, denominada algunas veces **tabla ANOVA**. Con la información de la tabla 5.3, consideremos ahora la siguiente variable:

$$
\begin{aligned}
F &= \frac{\text{SCP de SCE}}{\text{SCP de SRC}} \\
&= \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum \hat{u}_i^2 / (n-2)} \\
&= \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2}
\end{aligned} \tag{5.9.1}
$$

Si suponemos que las perturbaciones $u_i$ están normalmente distribuidas, lo cual se cumple para el MCRLN, y si la hipótesis nula ($H_0$) es que $\beta_2 = 0$, puede demostrarse que la variable $F$ de la ecuación (5.9.1) satisface la distribución $F$ con 1 gl en el numerador y $(n-2)$gl en el denominador. (Véase la prueba en el apéndice 5A, sección 5A.3. Las propiedades generales de la distribución $F$ se explican en el **apéndice A**.)

¿Qué uso puede hacerse de la razón $F$ anterior? Puede demostrarse[18] que

$$E\left(\hat{\beta}_2^2 \sum x_i^2\right) = \sigma^2 + \beta_2^2 \sum x_i^2 \tag{5.9.2}$$

y

$$E \frac{\sum \hat{u}_i^2}{n-2} = E(\hat{\sigma}^2) = \sigma^2 \tag{5.9.3}$$

(Observe que $\beta_2$ y $\sigma^2$ al lado derecho de estas ecuaciones son los verdaderos parámetros.) Por tanto, si $\beta_2$ es en realidad cero, ambas ecuaciones (5.9.2) y (5.9.3) proporcionan estimaciones idénticas del verdadero $\sigma^2$. En esta situación, la variable explicativa $X$ no tiene influencia lineal alguna sobre $Y$, y toda la variación en $Y$ se explica con las perturbaciones aleatorias $u_i$. Por otra parte, si $\beta_2$ es diferente de cero, (5.9.2) y (5.9.3) serán diferentes y parte de la variación en $Y$ se atribuirá a $X$. Por consiguiente, la razón $F$ de (5.9.1) constituye una prueba de la hipótesis nula $H_0$: $\beta_2 = 0$. Como todas las cantidades que forman parte de esta ecuación se obtienen de la muestra disponible, esta razón $F$ constituye un estadístico de prueba para verificar la hipótesis nula de que el verdadero $\beta_2$ es igual a cero. Sólo debe calcularse la razón $F$ y compararla con el valor crítico $F$ obtenido de las tablas $F$ en el nivel de significancia seleccionado, u obtener el **valor *p*** del estadístico $F$ calculado.

[18] Para una demostración, véase K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, Nueva York, 1960, pp. 278-280.

**TABLA 5.4**
**Tabla ANOVA para el ejemplo de los salarios y el nivel de escolaridad**

| Fuente de variación | SC | gl | SCP | |
|---|---|---|---|---|
| Debido a la regresión (SCE) | 95.4255 | 1 | 95.4255 | $F = \frac{95.4255}{0.8811}$ |
| Debido a los residuos (SCR) | 9.6928 | 11 | 0.8811 | $= 108.3026$ |
| SCT | 105.1183 | 12 | | |

Para esclarecer esto, continuamos con el ejemplo ilustrativo. La tabla ANOVA para este ejemplo se presenta en la tabla 5.4. El valor $F$ calculado es 108.3026. El valor $p$ de este estadístico $F$ correspondiente a 1 y 8 gl no puede obtenerse de la tabla $F$ dada en el **apéndice D,** pero con tablas estadísticas electrónicas se demuestra que el valor $p$ es 0.0000001, una probabilidad en efecto muy pequeña. Si escoge el método del nivel de significancia para la prueba de hipótesis y fija $\alpha$ en 0.01, o en un nivel de 1%, se verá que la $F$ calculada de 108.3026 es obviamente significativa en este nivel. Por tanto, si rechazamos la hipótesis nula de que $\beta_2 = 0$, la probabilidad de cometer un error tipo I es muy pequeña. Para todo fin práctico, la muestra no pudo provenir de una población con un valor $\beta_2$ igual a cero, y se puede concluir con gran confianza que $X$, la educación, sí afecta $Y$, el salario promedio.

Consulte el teorema 5.7 del apéndice 5A.1, que establece que el cuadrado del valor $t$ con $k$ gl es un valor $F$ con 1 gl en el numerador y $k$ gl en el denominador. En nuestro ejemplo, si suponemos que $H_0$: $\beta_2 = 0$, entonces con (5.3.2) se verifica fácilmente que el valor $t$ estimado es 10.41. Este valor $t$ tiene 11 gl. Según la misma hipótesis nula, el valor $F$ era 108.3026 con 1 y 11 gl. De donde $(10.3428)^2 =$ valor $F$, excepto por errores de redondeo.

Así, las pruebas $t$ y $F$ proporcionan dos formas alternas, pero complementarias, de probar la hipótesis nula de que $\beta_2 = 0$. Si éste es el caso, ¿por qué no sólo confiar en la prueba $t$ y no preocuparse por la prueba $F$ y el análisis de varianza que la acompaña? Para el modelo con dos variables, en realidad no es necesario recurrir a la prueba $F$. Pero cuando consideremos el tema de la regresión múltiple, veremos que la prueba $F$ tiene diversas aplicaciones interesantes que la hacen un método muy útil y eficaz para demostrar hipótesis estadísticas.

## 5.10 Aplicación del análisis de regresión: problema de predicción

Con base en los datos muestrales de la tabla 3.2 se obtuvo la siguiente regresión muestral:

$$\hat{Y}_i = -0.0144 + 0.7240X_i \tag{3.6.2}$$

donde $\hat{Y}_i$ es el estimador del verdadero $E(Y_i)$ correspondiente a $X$ dada. ¿De qué sirve esta **regresión histórica**? Para "predecir" o "pronosticar" el salario promedio futuro $Y$ correspondiente a algún nivel dado de escolaridad $X$. Ahora, hay dos clases de predicciones: 1) la predicción del valor de la media condicional de $Y$ correspondiente a un valor escogido $X$, por ejemplo, $X_0$, que es el punto sobre la línea de regresión poblacional misma (véase la figura 2.2), y 2) la predicción de un valor individual $Y$ correspondiente a $X_0$. Estas dos predicciones se llaman **predicción media** y **predicción individual**.

## Predicción media[19]

Para ordenar las ideas, suponga que $X_0 = 20$ y deseamos predecir $E(Y|X_0 = 20)$. Ahora, puede demostrarse que la regresión histórica (3.6.2) proporciona la estimación puntual de esta predicción media de la siguiente forma:

$$\begin{aligned} \hat{Y}_0 &= \hat{\beta}_1 + \hat{\beta}_2 X_0 \\ &= -0.0144 + 0.7240(20) \\ &= 14.4656 \end{aligned} \tag{5.10.1}$$

donde $\hat{Y}_0$ = estimador de $E(Y|X_0)$. Puede comprobarse que este predictor puntual es el mejor estimador lineal e insesgado (MELI).

Como $\hat{Y}_0$ es un estimador, es probable que éste sea diferente de su verdadero valor. La diferencia entre los dos valores dará alguna idea del error de predicción o pronóstico. Para evaluar este error es necesario encontrar la distribución muestral de $\hat{Y}_0$. En el apéndice 5A, sección 5A.4, se demuestra que en la ecuación (5.10.1) $\hat{Y}_0$ está normalmente distribuida con media $(\beta_1 + \beta_2 X_0)$ y una varianza dada por la siguiente fórmula:

$$\operatorname{var}(\hat{Y}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \tag{5.10.2}$$

Al reemplazar la $\sigma^2$ desconocida por su estimador insesgado $\hat{\sigma}^2$, vemos que la variable

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\operatorname{ee}(\hat{Y}_0)} \tag{5.10.3}$$

sigue una distribución $t$ con $n - 2$ gl. La distribución $t$ sirve por consiguiente para construir intervalos de confianza para el verdadero $E(Y_0 \mid X_0)$ y pruebas de hipótesis acerca de tal valor de la manera usual, a saber,

$$\Pr [\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2} \operatorname{ee}(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2} \operatorname{ee}(\hat{Y}_0)] = 1 - \alpha \tag{5.10.4}$$

donde ee $(\hat{Y}_0)$ se obtiene de (5.10.2).

Para nuestros datos (véase la tabla 3.2),

$$\begin{aligned} \operatorname{var}(\hat{Y}_0) &= 0.8936 \left[ \frac{1}{13} + \frac{(20 - 12)^2}{182} \right] \\ &= 0.3826 \end{aligned}$$

y

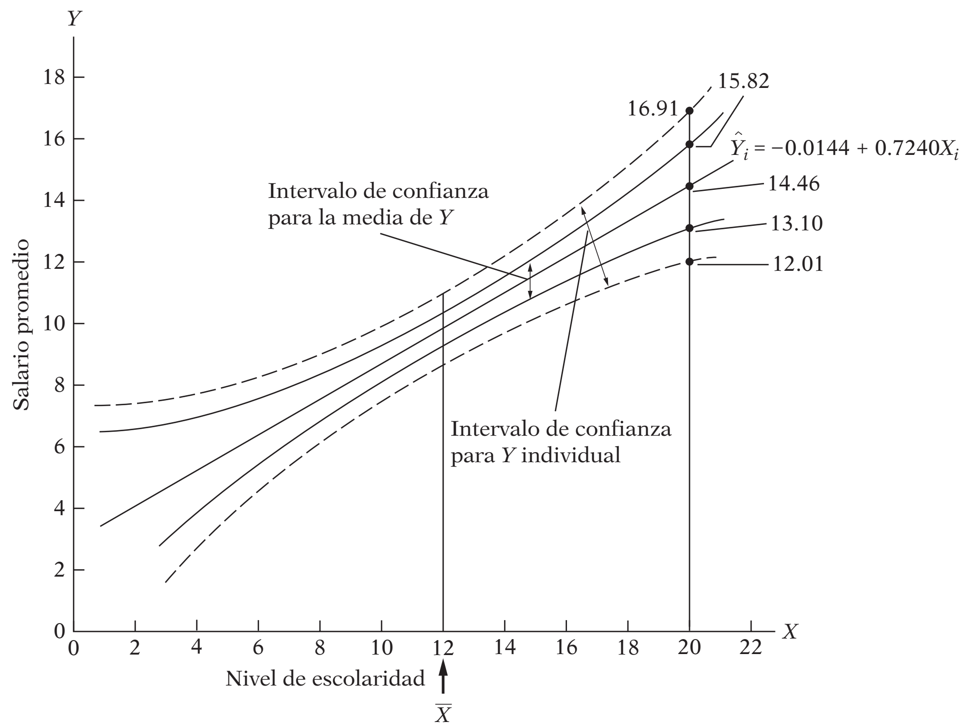
$$\operatorname{ee}(\hat{Y}_0) = 0.6185$$

Por tanto, el intervalo de confianza a 95% para el verdadero $E(Y|X_0) = \beta_1 + \beta_2 X_0$ está dado por

$$14.4656 - 2.201(.6185) \leq E(Y_0 \mid X = 20) \leq 14.4656 = 2.20(0.6185)$$

[19] Para las pruebas de las distintas afirmaciones, véase el apéndice 5A, sección 5A.4.

**FIGURA 5.6**
Intervalos (bandas) de confianza para la media de *Y* y los valores individuales de *Y*.




es decir,

$$13.1043 \leq E(Y \mid X = 20) \leq 15.8260 \tag{5.10.5}$$

Así, dado $X_0 = 100$, en muestreo repetido, en 95 de cada 100 intervalos como (5.10.5) estará el verdadero valor medio; la mejor estimación del verdadero valor medio es, por supuesto, la estimación puntual 75.3645.

Si obtenemos intervalos de confianza a 95% como (5.10.5) por cada valor de $X$ en la tabla 3.2, se obtiene lo que se conoce como **intervalo de confianza** o **banda de confianza**, para la función de regresión poblacional, que se presenta en la figura 5.6.

## Predicción individual

Si lo que interesa es predecir un valor individual $Y$, $Y_0$ correspondiente a un valor dado de $X$, digamos, $X_0$, entonces, como se muestra en el apéndice 5, sección 5A.4, el mejor estimador lineal insesgado de $Y_0$ está dado también por (5.10.1), pero su varianza es la siguiente:

$$\operatorname{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \tag{5.10.6}$$

Además, se demuestra que $Y_0$ también sigue una distribución normal con media y varianza dadas por (5.10.1) y (5.10.6), respectivamente. Al sustituir $\hat{\sigma}^2$ por la desconocida $\sigma^2$, se colige que

$$t = \frac{Y_0 - \hat{Y}_0}{\operatorname{ee}(Y_0 - \hat{Y}_0)}$$

también sigue una distribución $t$. Por consiguiente, la distribución $t$ sirve para inferir sobre la verdadera $Y_0$. Al continuar con el ejemplo, vemos que la predicción puntual de $Y_0$ es 14.4656, igual a $\hat{Y}_0$, y su varianza es 1.2357 (el lector debe verificar este cálculo). Por tanto, el intervalo de confianza a 95% para $Y_0$ correspondiente a $X_0 = 100$ es

$$(12.0190 \le Y_0 \mid X_0 = 20 \le 16.9122) \tag{5.10.7}$$

Si comparamos este intervalo con (5.10.5), vemos que el intervalo de confianza para la $Y_0$ individual es más amplio que el intervalo para el valor medio de $Y_0$. (¿Por qué?) Calculamos los intervalos de confianza como en (5.10.7) condicionales a los valores de $X$ dados en la tabla 3.2 y se obtiene la banda de confianza a 95% para los valores individuales $Y$ correspondientes a estos valores de $X$. Esta banda de confianza, al igual que la banda de confianza para $\hat{Y}_0$ asociada con los mismos $X$, se muestran en la figura 5.6.

Note una característica importante de las bandas de confianza de la figura 5.6. La amplitud más pequeña de estas bandas se presenta cuando $X_0 = \bar{X}$. (¿Por qué?) Sin embargo, aumenta considerablemente a medida que $X_0$ se aleja de $\bar{X}$. (¿Por qué?) Este cambio indicaría que la capacidad de predicción de la línea de regresión muestral *histórica* decrece mucho a medida que $X_0$ se aleja cada vez más de $\bar{X}$. **Por consiguiente, se debe tener cuidado al "extrapolar" la línea de regresión histórica para predecir $E(Y \mid X_0)$ o $Y_0$ asociada a una $X_0$ dada muy alejada de la media muestral $\bar{X}$.**

## 5.11 Informe de resultados del análisis de regresión

Hay diversas formas de presentar los resultados de un análisis de regresión; sin embargo, en este texto utilizaremos el siguiente formato, con el ejemplo de los salarios y el nivel de escolaridad del capítulo 3 a manera de ilustración:

$$\begin{aligned}
\hat{Y}_i &= -0.0144 \quad + \quad 0.7240X_i \\
\text{ee} &= (0.9317) \qquad (0.0700) \qquad r^2 = 0.9065 \\
t &= (-0.0154) \qquad (10.3428) \qquad \text{gl} = 11 \\
p &= (0.987) \qquad (0.000) \qquad F_{1,11} = 108.30
\end{aligned} \tag{5.11.1}$$

En la ecuación (5.11.1), las cifras del primer conjunto de paréntesis son los errores estándar estimados de los coeficientes de regresión; las cifras del segundo conjunto son los valores $t$ estimados calculados de (5.3.2) según la hipótesis nula de que el verdadero valor poblacional de cada coeficiente de regresión individual es cero (es decir, $10.3428 = \frac{0.7240}{0.0700}$); y las cifras del tercer grupo son los valores $p$ estimados. Por tanto, para 11 gl, la probabilidad de obtener un valor $t$ igual o mayor que 10.3428 es de 0.00009, o prácticamente cero.

Al presentar los valores $p$ de los coeficientes $t$ estimados, se aprecia de inmediato el nivel exacto de significancia de cada valor $t$ estimado. Así, conforme a la hipótesis nula de que el verdadero valor de la pendiente poblacional es cero (es decir, que el nivel de escolaridad no produce ningún efecto en el salario promedio), la probabilidad exacta de obtener un valor $t$ igual o mayor que 10.3428 es prácticamente cero. Recuerde que cuanto menor sea el valor $p$, menor será también la probabilidad de cometer un error si se rechaza la hipótesis nula.

Ya mostramos la conexión entre los estadísticos $F$ y $t$, a saber, $F_{1,k} = t_k^2$. Según la hipótesis nula de que el verdadero $\beta_2 = 0$, la ecuación (5.11.1) muestra que el valor $F$ es 108.30 (para 1 gl en el numerador y 11 gl en el denominador), y el valor $t$ es cercano a 10.34 (11 gl); como se esperaba, el primer valor es igual al último valor elevado al cuadrado, salvo por errores de aproximación. Ya se analizó la tabla ANOVA para este problema.

## 5.12 Evaluación de los resultados del análisis de regresión

En la figura I.4 de la Introducción esbozamos la anatomía de la creación de modelos econométricos. Ahora que presentamos los resultados del análisis de regresión del ejemplo de salarios y niveles de escolaridad en (5.11.1) cabe cuestionar la bondad del modelo ajustado. ¿Qué tan "bueno" es el modelo ajustado? Necesitamos ciertos criterios para responder esta pregunta.

Primero, ¿están los signos de los coeficientes estimados de acuerdo con las expectativas teóricas o previas? *A priori*, $\beta_2$ en el ejemplo de los salarios y el nivel de escolaridad debe ser positivo. En el presente ejemplo, lo es. Segundo, si la teoría sostiene que la relación no debe ser sólo positiva sino también estadísticamente significativa, ¿es el caso en la presente aplicación? Como analizamos en la sección 5.11, el coeficiente del nivel de escolaridad no sólo es positivo, sino también estadísticamente significativo, es decir, diferente de cero; el valor $p$ del valor $t$ estimado es muy pequeño. Valen los mismos comentarios para el coeficiente del intercepto. Tercero, ¿qué tan bien explica el modelo de regresión la variación en el ejemplo? Se puede responder con $r^2$. En nuestro ejemplo, $r^2$ es de alrededor de 0.90, un valor muy alto si consideramos que $r^2$ puede ser máximo 1.

Así, parece muy bueno el modelo escogido para explicar el comportamiento de los salarios promedio. Pero antes de comprometerse con él, sería interesante averiguar si satisface los supuestos del MCRLN. No veremos ahora los diversos supuestos, pues la simplicidad del modelo es clara. Sólo hay un supuesto que se puede verificar, a saber, el de normalidad del término de perturbación, $u_i$. Recuerde que las pruebas $t$ y $F$ requieren que el término de error siga una distribución normal. De lo contrario, el procedimiento de prueba no será válido en muestras pequeñas, o finitas.

### Pruebas de normalidad

Aunque se han estudiado diversas pruebas de normalidad en la teoría, sólo consideraremos tres: 1) histograma de residuos, 2) gráfica de probabilidad normal (GPN) y 3) prueba **Jarque-Bera**.

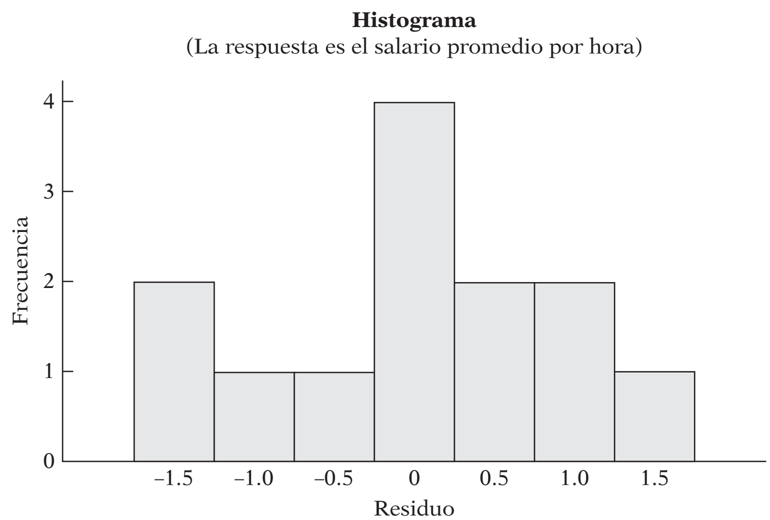
#### *Histograma de residuos*

Es un simple dispositivo gráfico para saber algo sobre la forma de la función de densidad poblacional (FDP) de una variable aleatoria. En el eje horizontal se dividen los valores de la variable de interés (por ejemplo, los residuos de MCO) en intervalos convenientes, y sobre cada intervalo de clase se construyen rectángulos cuya altura sea igual al número de observaciones (es decir, la frecuencia) para ese intervalo de clase. Si mentalmente se coloca la curva de distribución normal en forma de campana sobre el histograma, se tendrá cierta idea sobre la pertinencia o no de la aproximación normal (FDP). En la figura 5.7 se presenta el histograma de residuos correspondiente a la regresión de salarios y nivel de escolaridad.

Este diagrama muestra que los residuos no tienen distribución normal perfecta; para una variable distribuida normalmente, la asimetría (una medida de la simetría) debe ser cero, y la curtosis (que mide si la distribución normal es alta o baja), 3.

Siempre es aconsejable trazar el histograma de los residuos de cualquier regresión como método aproximado y rápido para probar el supuesto de normalidad.

**FIGURA 5.7**
Histograma de residuos de los datos de salarios y nivel de escolaridad.




### *Gráfica de probabilidad normal*

Un dispositivo gráfico relativamente sencillo para estudiar la forma de la función de densidad de probabilidad (FDP) de una variable aleatoria es la **gráfica de probabilidad normal (GPN)**, la cual utiliza el *papel de probabilidad normal*, especialmente diseñado para gráficas. Sobre el eje horizontal, o eje *X*, se grafican los valores de la variable de interés (por ejemplo, los residuos de MCO, $\hat{u}_i$), y sobre el eje vertical, o eje *Y*, el valor esperado de esta variable si estuviera normalmente distribuida. Por tanto, si la variable fuese de la población normal, la GPN sería más o menos una línea recta. La GPN de los residuos correspondientes a la regresión de los salarios y el nivel de escolaridad se muestra en la figura 5.8, que se obtuvo con el software MINITAB, versión 15. Como se observó antes, si la línea resultante en la GPN es aproximadamente una línea recta, se puede concluir que la variable de interés está normalmente distribuida. En la figura 5.7 se aprecia que los residuos del ejemplo están aproximadamente distribuidos normalmente, pues al parecer una línea recta se ajusta a los datos de forma razonable.

Con el MINITAB también se obtiene la **prueba de normalidad Anderson-Darling**, conocida como **estadístico $A^2$**. La hipótesis nula es que la variable en cuestión está normalmente distribuida. Como se muestra en la figura 5.8, para el ejemplo, el estadístico $A^2$ calculado es 0.289. El *valor p* de obtener tal valor de $A^2$ es de 0.558, razonablemente alto. Por consiguiente, no rechazamos la hipótesis de que los residuos del ejemplo ilustrativo están normalmente distribuidos. A propósito, en la figura 5.8 se muestran los parámetros de la distribución (normal), la media es aproximadamente 0, y la desviación estándar, de casi 0.8987.

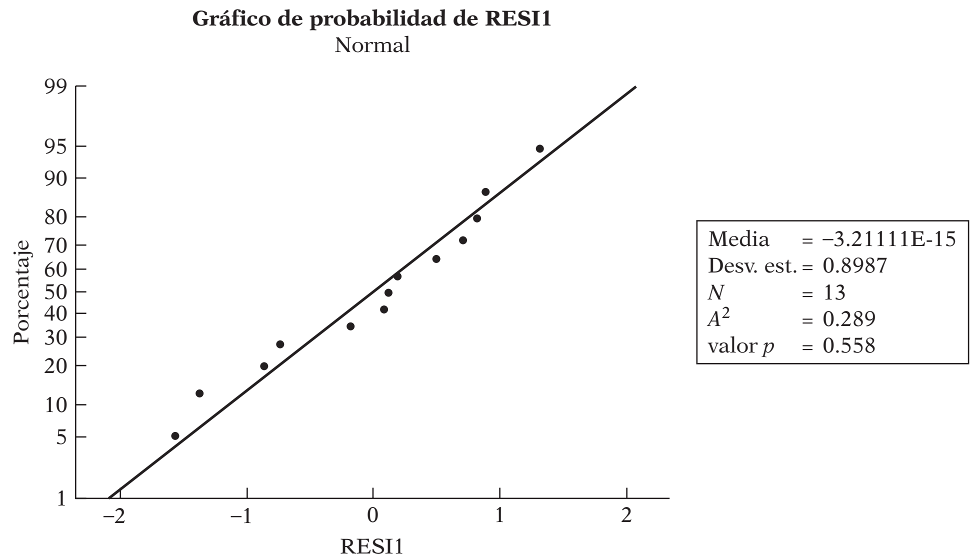
### *Prueba de normalidad de Jarque-Bera (JB)*[20]

La prueba de normalidad JB es una prueba *asintótica*, o de muestras grandes. También se basa en los residuos de MCO. Esta prueba calcula primero la **asimetría** y la **curtosis** (analizadas en el **apéndice A**) de los residuos de MCO, con el siguiente estadístico de prueba:

$$\text{JB} = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \qquad \textbf{(5.12.1)}$$

[20] Véase C.M. Jarque y A.K. Bera, "A Test for Normality of Observations and Regression Residuals", *International Statistical Review*, vol. 55, 1987, pp. 163-172.

**FIGURA 5.8**
Residuos de la regresión de salarios y nivel de escolaridad.




donde $n$ = tamaño de la muestra, $S$ = coeficiente de asimetría y $K$ = coeficiente de curtosis. Para una variable normalmente distribuida, $S = 0$ y $K = 3$. Por tanto, la prueba de normalidad JB constituye una prueba de la hipótesis conjunta de que $S$ y $K$ son 0 y 3, respectivamente. En este caso, se espera que el valor del estadístico JB sea igual a cero.

De acuerdo con la hipótesis nula, la cual afirma que los residuos están normalmente distribuidos, Jarque y Bera mostraron que *asintóticamente (es decir, en muestras grandes) el estadístico JB dado en (5.12.1) sigue la distribución ji cuadrada, con 2 gl.* Si el valor $p$ calculado del estadístico JB es lo bastante bajo en una aplicación, lo cual sucederá si el valor del estadístico difiere en gran medida de cero, se puede rechazar la hipótesis de que los residuos están normalmente distribuidos. Pero si el valor $p$ es razonablemente alto, lo cual sucede cuando el valor del estadístico está cerca de cero, no rechazamos la suposición de normalidad.

En nuestro ejemplo, el estadístico JB estimado para la regresión de salarios y nivel de educación es 0.8286. La hipótesis nula de que los residuos en el presente ejemplo están distribuidos normalmente no puede rechazarse, pues el valor $p$ de obtener un estadístico JB igual o mayor que 0.8286 es de alrededor de 0.66, o 66%. Esta probabilidad es muy alta. Observe que, aunque la regresión tiene 13 observaciones, éstas se obtuvieron de una muestra de 528 observaciones, lo que parece razonablemente alto.

## Otras pruebas del ajuste del modelo

Recuerde que el MCRLN tiene muchos supuestos adicionales al de la normalidad del término de error. A medida que examinemos la teoría econométrica, consideraremos diversas pruebas de la bondad del modelo (véase el capítulo 13). Hasta entonces, recuerde que la elaboración de modelos de regresión se basa en diversos supuestos simplificadores que quizá no sean válidos en todos los casos.

***Un ejemplo para concluir***

Regresemos al ejemplo 3.2 respecto del gasto alimentario en India. Con los datos de (3.7.2) y el formato (5.11.1), obtenemos la siguiente ecuación para el gasto:

$$
\begin{aligned}
\widehat{\text{GasAli}}_i &= 94.2087 \;+\; 0.4368\ \text{GasTot}_i \\
\text{ee} &= (50.8563) \quad (0.0783) \\
t &= (1.8524) \quad (5.5770) \\
p &= (0.0695) \quad (0.0000)^* \\
r^2 &= 0.3698; \quad \text{gl} = 53 \\
F_{1,53} &= 31.1034 \quad (\text{valor } p = 0.0000)^*
\end{aligned}
\qquad \textbf{(5.12.2)}
$$

donde * significa extremadamente pequeño.

En primer lugar, interpretemos esta regresión. Como se esperaba, hay una relación positiva entre el gasto alimentario y el gasto total. Si este último se incrementara una rupia, en promedio, el gasto en comida aumentaría casi 44 paisas. Si el gasto total fuera nulo, el gasto promedio en comida sería más o menos de 94 rupias. Por supuesto, esta interpretación mecánica del intercepto quizá no tenga mucho sentido en la economía. El valor $r^2$ de casi 0.37 significa que 37% de la variación en el gasto alimentario se explica por el gasto total, una aproximación para el ingreso.

Suponga que deseamos probar la hipótesis nula de que no existe relación entre el gasto alimentario y el total; es decir, el verdadero coeficiente de la pendiente $\beta_2 = 0$. El valor estimado de $\beta_2$ es 0.4368. Si la hipótesis nula es cierta, ¿cuál es la probabilidad de obtener un valor igual a 0.4368? Según la hipótesis nula, se observa de (5.12.2) que el valor $t$ es 5.5770 y que el valor $p$ de obtener dicho valor $t$ es prácticamente cero. En otras palabras, se puede rechazar la hipótesis nula con toda justificación. Pero suponga que la hipótesis nula fuese que $\beta_2 = 0.5$, ¿qué pasaría? Con la prueba $t$ obtenemos

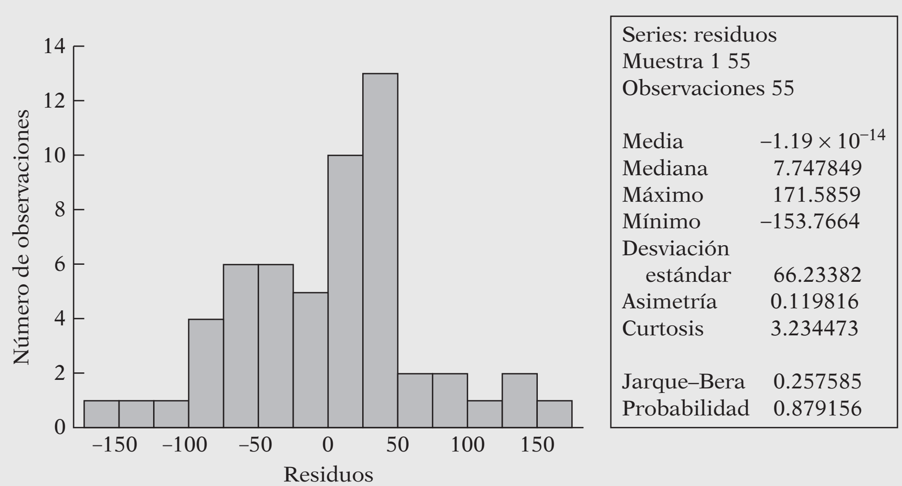
$$t = \frac{0.4368 - 0.5}{0.0783} = -0.8071$$

La probabilidad de obtener una $|t|$ de 0.8071 es mayor que 20%. Por tanto, no se rechaza la hipótesis de que el verdadero valor de $\beta_2$ sea 0.5.

Observe que, conforme a la hipótesis nula, el verdadero coeficiente de la pendiente es cero, el valor $F$ es 31.1034, como se muestra en (5.12.2). Según la misma hipótesis nula, se obtiene un valor $t$ de 5.5770. Si elevamos al cuadrado este valor, se obtiene 31.1029, que es casi el mismo que el valor $F$, con lo cual se muestra de nuevo la estrecha relación entre $t$ y el estadístico $F$. (*Nota:* el número de gl del numerador del estadístico $F$ debe ser 1, lo cual ocurre en este caso.)

Con los residuos estimados de la regresión, ¿qué podemos decir respecto de la distribución de probabilidad del término de error? La respuesta se da en la figura 5.9. Como ahí se muestra,

**FIGURA 5.9**
Residuos de la regresión del gasto alimentario.




(*continúa*)

*Un ejemplo para concluir* (*continuación*)

los residuos de la regresión del gasto alimentario parecen estar simétricamente distribuidos. Al aplicarse la prueba Jarque-Bera, se hace evidente que el estadístico JB es casi de 0.2576, y que la probabilidad de obtener tal estadístico con el supuesto de normalidad es más o menos de 88%. En consecuencia, no rechazamos la hipótesis de que los términos de error están normalmente distribuidos. Pero hay que tener presente que el tamaño de la muestra de 55 observaciones tal vez no sea lo bastante grande.

Dejamos al lector la tarea de establecer los intervalos de confianza para los dos coeficientes de regresión, obtener la gráfica de probabilidad normal, y realizar predicciones individuales y para la media.

## Resumen y conclusiones

1. La estimación y las pruebas de hipótesis constituyen las dos ramas principales de la estadística clásica. Después de analizar el problema de la estimación en los capítulos 3 y 4, en este capítulo contemplamos las pruebas de hipótesis.
2. Las pruebas de hipótesis responden a esta pregunta: ¿un determinado hallazgo es compatible con la hipótesis planteada o no lo es?
3. Hay dos enfoques mutuamente complementarios para responder a la pregunta anterior: **intervalos de confianza** y **pruebas de significancia**.
4. El método de intervalos de confianza se basa en el concepto de **estimación por intervalos**. Un estimador por intervalos es un intervalo o rango construido de manera que tiene una probabilidad específica de contener dentro de sus límites al verdadero valor del parámetro desconocido. El intervalo así construido se conoce como **intervalo de confianza**, que suele definirse en forma porcentual, como 90 o 95%. El intervalo de confianza proporciona un conjunto de hipótesis factibles acerca del valor del parámetro desconocido. Si el valor al que se refiere la hipótesis nula se encuentra en el intervalo de confianza, la hipótesis no se rechaza, y si se encuentra fuera de este intervalo, la hipótesis nula puede rechazarse.
5. En el procedimiento de **pruebas de significancia** se elabora un **estadístico de prueba** y se examina su distribución muestral según la hipótesis nula. El estadístico de prueba sigue con frecuencia una distribución de probabilidad conocida como normal, $t$, $F$, o ji cuadrada. Una vez calculado el estadístico de prueba (por ejemplo, el estadístico $t$) a partir de los datos disponibles, resulta sencillo obtener su valor $p$. El valor $p$ da la probabilidad exacta de obtener el estadístico de prueba estimado conforme a la hipótesis nula. Si este valor $p$ es pequeño, se rechaza la hipótesis nula, pero si es grande puede no rechazarse. El investigador debe decidir qué constituye un valor $p$ pequeño o grande. Al seleccionar el valor $p$, el investigador debe considerar la probabilidad de cometer **errores tipo I** y **tipo II**.
6. En la práctica, se debe tener cuidado al determinar $\alpha$, la probabilidad de cometer un **error tipo I**, al asignarle valores arbitrarios, como 1, 5 o 10%. Es mejor hacer referencia al **valor *p*** del estadístico de prueba. Además, la significancia estadística de una estimación no debe confundirse con su significancia práctica.
7. Por supuesto, las pruebas de hipótesis suponen que el modelo seleccionado para el análisis empírico es adecuado en el sentido de que no viola alguno(s) del (los) supuestos(s) que sirven de base al modelo clásico de regresión lineal normal. Por tanto, las pruebas sobre la bondad del modelo deben preceder a las pruebas de hipótesis. En este capítulo describimos una de estas pruebas, la **prueba de normalidad**, cuyo fin es establecer si el término de error sigue una distribución normal. Como en muestras pequeñas, o finitas, las pruebas $t$, $F$ y ji cuadrada requieren el supuesto de normalidad, es importante verificar formalmente este supuesto.
8. Si el modelo se considera correcto o adecuado, puede emplearse para pronosticar. Pero al pronosticar los valores futuros del regresando o variable dependiente, no debe salirse demasiado del rango muestral de los valores de la regresora, o variable explicativa. De lo contrario, los errores de pronóstico pueden aumentar en forma drástica.

## EJERCICIOS

### Preguntas

5.1. Establezca si las siguientes afirmaciones son verdaderas, falsas o inciertas. Sea preciso.

*a*) La prueba $t$ de significancia estudiada en este capítulo requiere que las distribuciones muestrales de los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ sigan una distribución normal.

*b*) Aunque el término de perturbación en el MCRL no esté normalmente distribuido, los estimadores de MCO continúan siendo insesgados.

*c*) Si no hay intercepto en el modelo de regresión, las $u_i (= \hat{u}_i)$ estimadas no sumarán cero.

*d*) El valor $p$ y el tamaño de un estadístico de prueba tienen el mismo significado.

*e*) En un modelo de regresión que contenga el intercepto, la suma de los residuos es siempre cero.

*f*) Si no se rechaza una hipótesis nula, es verdadera.

*g*) Entre mayor sea el valor de $\sigma^2$, mayor será la varianza de $\hat{\beta}_2$ dada en (3.3.1).

*h*) Las medias condicional e incondicional de una variable aleatoria significan lo mismo.

*i*) En una FRP de dos variables, si el coeficiente de la pendiente $\beta_2$ es cero, el intercepto $\beta_1$ se estima por la media muestral $\bar{Y}$.

*j*) La varianza condicional, $\text{var}(Y_i \mid X_i) = \sigma^2$, y la varianza incondicional de $Y$, $\text{var}(Y) = \sigma_Y^2$, serían la misma si $X$ no tuviera influencia en $Y$.

5.2. Construya la tabla ANOVA como la de la tabla 5.4 para el modelo de regresión dado en (3.7.2) y pruebe la hipótesis de que no existe relación entre el gasto en alimentos y el gasto total en India.

5.3. Consulte la regresión de la demanda de teléfonos celulares de la ecuación (3.7.3).

*a*) ¿El coeficiente estimado del intercepto es significativo en el nivel de significancia de 5%? ¿Qué hipótesis nula está probando?

*b*) ¿El coeficiente estimado de la pendiente es significativo en el nivel de significancia de 5%? ¿En qué hipótesis nula se basa?

*c*) Establezca un intervalo de confianza a 95% para el verdadero coeficiente de la pendiente.

*d*) ¿Cual es el valor pronosticado de la media de los teléfonos celulares demandados si el ingreso *per cápita* es de 9 000 dólares? ¿Cuál es el intervalo de confianza a 95% para el valor pronosticado?

5.4. Sea $\rho^2$ el verdadero coeficiente de determinación poblacional. Suponga que desea probar la hipótesis de que $\rho^2 = 0$. Explique verbalmente cómo probar esta hipótesis. *Sugerencia:* Utilice la ecuación (3.5.11). Véase también el ejercicio 5.7.

5.5. Lo que se conoce como la **línea característica** del análisis de inversión moderno es sencillamente la línea de regresión obtenida del siguiente modelo:

$$r_{it} = \alpha_i + \beta_i r_{mt} + u_t$$

donde $r_{it}$ = la tasa de rendimiento del $i$-ésimo valor en el tiempo $t$

$r_{mt}$ = la tasa de rendimiento del portafolio del mercado en el tiempo $t$

$u_t$ = término de perturbación estocástica

En este modelo, $\beta_i$ se conoce como **coeficiente beta** del $i$-ésimo valor, una medida del riesgo del mercado (o sistemático) de un valor.*

* Véase Haim Levy y Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, Nueva Jersey, 1984, capítulo 12.

Con base en 240 tasas de rendimiento mensuales para el periodo 1956-1976, Fogler y Ganapathy obtuvieron la siguiente línea característica para las acciones de IBM en relación con el índice de portafolio del mercado elaborado en la Universidad de Chicago:*

$$\hat{r}_{it} = 0.7264 + 1.0598 r_{mt} \qquad r^2 = 0.4710$$
$$\text{ee} = (0.3001)\ (0.0728) \qquad \text{gl} = 238$$
$$F_{1,238} = 211.896$$

*a*) Se dice que un valor cuyo coeficiente beta es mayor que uno es un valor volátil o agresivo. ¿Fueron las acciones de IBM valores volátiles en el periodo que se estudia?

*b*) ¿Es el coeficiente del intercepto significativamente diferente de cero? Si lo es, ¿cuál es su interpretación práctica?

5.6. La ecuación (5.3.5) también se escribe como

$$\Pr\left[\hat{\beta}_2 - t_{\alpha/2}\text{ee}(\hat{\beta}_2) < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}\text{ee}(\hat{\beta}_2)\right] = 1 - \alpha$$

Es decir, la desigualdad débil ($\leq$) puede reemplazarse por la desigualdad fuerte ($<$). ¿Por qué?

5.7. R. A. Fisher derivó la distribución muestral del coeficiente de correlación definido en (3.5.13). Si se supone que las variables $X$ y $Y$ tienen una distribución normal conjunta, es decir, si provienen de una distribución normal bivariada (véase el apéndice 4A, ejercicio 4.1), entonces, según el supuesto de que el coeficiente de correlación poblacional $\rho$ es cero, es posible demostrar que $t = r\sqrt{n-2}/\sqrt{1-r^2}$ sigue la distribución $t$ de Student con $n - 2$ gl.** Demuestre que este valor $t$ es idéntico al valor $t$ dado en (5.3.2) con la hipótesis nula de que $\beta_2 = 0$. Por tanto, establezca que con la misma hipótesis nula $F = t^2$. (Véase la sección 5.9.)

5.8. Considere el siguiente resultado de una regresión:†

$$\hat{Y}_i = 0.2033 + 0.6560 X_t$$
$$\text{ee} = (0.0976)\ (0.1961)$$
$$r^2 = 0.397 \qquad \text{SCR} = 0.0544 \qquad \text{SCE} = 0.0358$$

donde $Y$ = tasa de participación de la fuerza laboral (TPFL) de las mujeres en 1972 y $X$ = TPFL de las mujeres en 1968. Los resultados de la regresión se obtuvieron de una muestra de 19 ciudades de Estados Unidos.

*a*) ¿Cómo interpreta esta regresión?

*b*) Pruebe la hipótesis $H_0\colon \beta_2 = 1$ contra $H_1\colon \beta_2 > 1$. ¿Qué prueba utilizaría? ¿Por qué? ¿Cuáles son los supuestos subyacentes de la(s) prueba(s) que eligió?

*c*) Suponga que la TPFL para 1968 fue de 0.58 (o 58%). Con base en los resultados de la regresión anteriores, ¿cuál es la TPFL media en 1972? Establezca un intervalo de confianza de 95% para la predicción de la media.

*d*) ¿Cómo probaría la hipótesis de que el término de error en la regresión sobre la población está normalmente distribuido? Muestre los cálculos necesarios.

---

*H. Russell Fogler y Sundaram Ganapathy, *Financial Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1982, p. 13.

**Si en realidad $\rho$ es cero, Fisher demostró que $r$ sigue la misma distribución $t$ siempre que $X$ o $Y$ estén normalmente distribuidas. Sin embargo, si $\rho$ no es igual a cero, ambas variables deben estar normalmente distribuidas. Véase R. L. Anderson y T. A. Bancroft, *Statistical Theory in Research*, McGraw-Hill, Nueva York, 1952, pp. 87-88.

† Adaptado de Samprit Chatterjee, Ali S. Hadi y Bertram Price, *Regression Analysis by Example*, 3a. ed., Wiley Interscience, Nueva York, 2000, pp. 46-47.

**TABLA 5.5**
**Salario promedio y gasto por alumno (dólares), 1985**

Fuente: National Education Association, como se publicó en *Albuquerque Tribune,* 7 de noviembre de 1986.

| Observación | Salario | Gasto | Observación | Salario | Gasto |
|---|---|---|---|---|---|
| 1 | 19 583 | 3 346 | 27 | 22 795 | 3 366 |
| 2 | 20 263 | 3 114 | 28 | 21 570 | 2 920 |
| 3 | 20 325 | 3 554 | 29 | 22 080 | 2 980 |
| 4 | 26 800 | 4 642 | 30 | 22 250 | 3 731 |
| 5 | 29 470 | 4 669 | 31 | 20 940 | 2 853 |
| 6 | 26 610 | 4 888 | 32 | 21 800 | 2 533 |
| 7 | 30 678 | 5 710 | 33 | 22 934 | 2 729 |
| 8 | 27 170 | 5 536 | 34 | 18 443 | 2 305 |
| 9 | 25 853 | 4 168 | 35 | 19 538 | 2 642 |
| 10 | 24 500 | 3 547 | 36 | 20 460 | 3 124 |
| 11 | 24 274 | 3 159 | 37 | 21 419 | 2 752 |
| 12 | 27 170 | 3 621 | 38 | 25 160 | 3 429 |
| 13 | 30 168 | 3 782 | 39 | 22 482 | 3 947 |
| 14 | 26 525 | 4 247 | 40 | 20 969 | 2 509 |
| 15 | 27 360 | 3 982 | 41 | 27 224 | 5 440 |
| 16 | 21 690 | 3 568 | 42 | 25 892 | 4 042 |
| 17 | 21 974 | 3 155 | 43 | 22 644 | 3 402 |
| 18 | 20 816 | 3 059 | 44 | 24 640 | 2 829 |
| 19 | 18 095 | 2 967 | 45 | 22 341 | 2 297 |
| 20 | 20 939 | 3 285 | 46 | 25 610 | 2 932 |
| 21 | 22 644 | 3 914 | 47 | 26 015 | 3 705 |
| 22 | 24 624 | 4 517 | 48 | 25 788 | 4 123 |
| 23 | 27 186 | 4 349 | 49 | 29 132 | 3 608 |
| 24 | 33 990 | 5 020 | 50 | 41 480 | 8 349 |
| 25 | 23 382 | 3 594 | 51 | 25 845 | 3 766 |
| 26 | 20 627 | 2 821 | | | |

## Ejercicios empíricos

5.9. La tabla 5.5 proporciona datos sobre el salario promedio de un maestro de escuela pública (el sueldo anual está en dólares) y el gasto en educación pública por alumno (dólares) para 1985 en los 50 estados y el Distrito de Columbia en Estados Unidos.

A fin de averiguar si existe alguna relación entre el salario del maestro y el gasto por alumno en las escuelas públicas, se sugirió el siguiente modelo: $\text{Sueldo}_i = \beta_1 + \beta_2 \text{Gasto}_i + u_i$, donde la variable Sueldo es el salario del maestro y la variable Gasto significa gasto por alumno.

*a*) Grafique los datos y trace la línea de regresión.

*b*) Suponga, con base en el inciso *a*), que decide estimar el modelo de regresión dado antes. Obtenga las estimaciones de los parámetros, sus errores estándar, $r^2$, la SCR y la SCE.

*c*) Interprete la regresión. ¿Tiene sentido económico?

*d*) Establezca un intervalo de confianza de 95% para $\beta_2$. ¿Rechazaría la hipótesis de que el verdadero coeficiente de la pendiente es 3.0?

*e*) Obtenga el valor individual pronosticado y la media del Sueldo, si el gasto por alumno es de 5 000 dólares. También establezca intervalos de confianza de 95% para la verdadera media y el verdadero valor individual del Sueldo, para la cifra correspondiente al gasto.

*f*) ¿Cómo probaría la suposición de la normalidad del término de error? Muestre la(s) prueba(s) utilizada(s).

5.10. Consulte el ejercicio 3.20 para construir las tablas ANOVA y probar la hipótesis de que no existe ninguna relación entre la productividad y la remuneración salarial real. Haga esto con el sector de negocios y con el no agrícola.

5.11. Consulte el ejercicio 1.7.

*a*) Grafique los datos sobre impresiones en el eje vertical y el gasto publicitario en el horizontal. ¿Qué tipo de relación observa?

*b*) ¿Sería apropiado ajustar un modelo de regresión de dos variables a los datos? ¿Por qué? Si la respuesta es negativa, ¿qué tipo de modelo de regresión se ajustaría a los datos? ¿Cuenta con las herramientas necesarias para ajustar dicho modelo?

*c*) Suponga que no se grafican los datos y que sólo ajusta el modelo de regresión con dos variables a los datos. Obtenga los resultados usuales de la regresión. Guarde dichos resultados para revisar después este problema.

5.12. Consulte el ejercicio 1.1.

*a*) Grafique el índice de precios al consumidor (IPC) estadounidense en relación con el canadiense. ¿Qué revela la gráfica?

*b*) Suponga que desea predecir el IPC de EU con base en el canadiense. Elabore un modelo apropiado.

*c*) Pruebe la hipótesis de que no existe relación entre ambos IPC. Utilice $\alpha = 5\%$. Si rechaza la hipótesis nula, ¿significa que el IPC canadiense "condiciona" al IPC estadounidense? ¿Por qué?

5.13. Consulte el ejercicio 3.22.

*a*) Estime las dos regresiones dadas en dicho ejercicio y obtenga los errores estándar, así como los demás resultados usuales.

*b*) Pruebe la hipótesis de que las perturbaciones en los dos modelos de regresión están normalmente distribuidas.

*c*) En la regresión del precio del oro, pruebe la hipótesis de que $\beta_2 = 1$; es decir, que existe una relación uno a uno entre los precios del oro y el IPC (el oro tiene una cobertura perfecta). ¿Cuál es el valor *p* del estadístico de prueba estimado?

*d*) Repita el paso *c*) para la regresión del índice de la Bolsa de Valores de Nueva York (NYSE). ¿La inversión en el mercado de valores representa una cobertura perfecta contra la inflación? ¿Cuál es la hipótesis nula que está probando? ¿Cuál es su valor *p*?

*e*) Entre el oro y el mercado de valores, ¿qué inversión elegiría? ¿En qué basa su decisión?

5.14. La tabla 5.6 presenta datos sobre el producto nacional bruto y cuatro definiciones de la oferta monetaria de Estados Unidos durante el periodo 1970-1983. Al hacer regresiones del PNB respecto de las diversas definiciones de dinero, se obtienen los resultados de la tabla 5.7:

Los monetaristas o partidarios de la teoría cuantitativa sostienen que el ingreso nominal (el PNB nominal) se determina en gran medida por cambios en la cantidad o en la reserva de dinero, aunque no hay consenso sobre la definición "correcta" de dinero. Con los resultados de la tabla anterior, considere las siguientes preguntas:

*a*) ¿Qué definición de oferta monetaria parece estrechamente relacionada con el PNB nominal?

*b*) Como los términos $r^2$ son uniformemente elevados, ¿significa esto que no importa la elección de la definición de dinero?

*c*) Si el Banco de la Reserva Federal desea controlar la oferta monetaria, ¿cuál de estas medidas de dinero es una mejor meta para ese propósito? ¿Puede deducir su respuesta de los resultados de la regresión?

5.15. Suponga que la ecuación de una **curva de indiferencia** entre dos bienes está dada por

$$X_i Y_i = \beta_1 + \beta_2 X_i$$

¿Cómo estimaría los parámetros de este modelo? Aplique el modelo anterior a los datos de la tabla 5.8 y comente sus resultados.

**TABLA 5.6**
**PNB y cuatro medidas de la oferta monetaria**

Fuente: *Economic Report of the President*, 1985, datos del PNB de la tabla B-1, p. 232; datos de la oferta monetaria de la tabla B-61, p. 303.

| Año | PNB miles de millones de dólares | Medida de oferta monetaria, miles de millones de dólares | | | |
|---|---|---|---|---|---|
| | | $M_1$ | $M_2$ | $M_3$ | L |
| 1970 | 992.70 | 216.6 | 628.2 | 677.5 | 816.3 |
| 1971 | 1 077.6 | 230.8 | 712.8 | 776.2 | 903.1 |
| 1972 | 1 185.9 | 252.0 | 805.2 | 886.0 | 1 023.0 |
| 1973 | 1 326.4 | 265.9 | 861.0 | 985.0 | 1 141.7 |
| 1974 | 1 434.2 | 277.6 | 908.5 | 1 070.5 | 1 249.3 |
| 1975 | 1 549.2 | 291.2 | 1 023.3 | 1 174.2 | 1 367.9 |
| 1976 | 1 718.0 | 310.4 | 1 163.6 | 1 311.9 | 1 516.6 |
| 1977 | 1 918.3 | 335.4 | 1 286.7 | 1 472.9 | 1 704.7 |
| 1978 | 2 163.9 | 363.1 | 1 389.1 | 1 647.1 | 1 910.6 |
| 1979 | 2 417.8 | 389.1 | 1 498.5 | 1 804.8 | 2 117.1 |
| 1980 | 2 631.7 | 414.9 | 1 632.6 | 1 990.0 | 2 326.2 |
| 1981 | 2 957.8 | 441.9 | 1 796.6 | 2 238.2 | 2 599.8 |
| 1982 | 3 069.3 | 480.5 | 1 965.4 | 2 462.5 | 2 870.8 |
| 1983 | 3 304.8 | 525.4 | 2 196.3 | 2 710.4 | 3 183.1 |

Definiciones:

$M_1$ = circulante + depósitos a la vista + cheques de viajero y otros depósitos a corto plazo (ODC).
$M_2$ = $M_1$ + transacciones de recompra (RC) entre bancos y eurodólares a corto plazo + saldos FMMM (fondos mutualistas del mercado monetario) + CAMM (cuentas de ahorro del mercado monetario) + ahorros y pequeños depósitos.
$M_3$ = $M_2$ + depósitos a largo plazo + transacciones de recompra a plazo (RC a término) + FMMM institucional.
L = $M_3$ + otros activos líquidos.

**TABLA 5.7**
**Regresiones PNB-oferta monetaria, 1970-1983**

1) $\widehat{\text{PNB}}_t = -787.4723 + 8.0863\, M_{1t}$ $\quad r^2 = 0.9912$
$\qquad (77.9664) \quad (0.2197)$

2) $\widehat{\text{PNB}}_t = -44.0626 + 1.5875\, M_{2t}$ $\quad r^2 = 0.9905$
$\qquad (61.0134) \quad (0.0448)$

3) $\widehat{\text{PNB}}_t = 159.1366 + 1.2034\, M_{3t}$ $\quad r^2 = 0.9943$
$\qquad (42.9882) \quad (0.0262)$

4) $\widehat{\text{PNB}}_t = 164.2071 + 1.0290\, L_t$ $\quad r^2 = 0.9938$
$\qquad (44.7658) \quad (0.0234)$

*Nota:* Las cifras entre paréntesis son los errores estándar estimados.

**TABLA 5.8**

| | | | | | |
|---|---|---|---|---|---|
| Consumo del bien *X*: | 1 | 2 | 3 | 4 | 5 |
| Consumo del bien *Y*: | 4 | 3.5 | 2.8 | 1.9 | 0.8 |

5.16. Desde 1986, *The Economist* publica el Índice Big Mac como medida burda y divertida para saber si las monedas internacionales se sitúan en su tipo de cambio "correcto", según la teoría de la **paridad del poder adquisitivo (PPA)**. La PPA sostiene que con una unidad de moneda debe ser posible comprar la misma canasta de bienes en todos los países. Los partidarios de la PPA argumentan que, a la larga, las monedas tienden a moverse hacia su PPA. *The Economist* utiliza la hamburguesa Big Mac de McDonald's como canasta representativa, y presenta la información en la tabla 5.9.

Considere el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

donde $Y$ = tipo de cambio del día y $X$ = PPA implícita del dólar.

*a*) Si se mantiene la PPA, ¿qué valores de $\beta_1$ y $\beta_2$ esperaría *a priori*?

## TABLA 5.9

**El estándar de la hamburguesa**

Fuente: McDonald's y *The Economist*, 1 de febrero de 2007.

| | Precios de la hamburguesa Big Mac | | | | |
|---|---|---|---|---|---|
| | En moneda local | En dólares | PPA* implícita del dólar | Tipo de cambio del día del dólar, 31 de enero | Sub (−)/ sobre (+) valuación de la moneda local frente al dólar, % |
| Arabia Saudita | Riyal 9.00 | 2.40 | 2.80 | 3.75 | −25 |
| Argentina | Peso 8.25 | 2.65 | 2.56 | 3.11 | −18 |
| Australia | A$3.45 | 2.67 | 1.07 | 1.29 | −17 |
| Brasil | Real 6.4 | 3.01 | 1.99 | 2.13 | −6 |
| Canadá | C$3.63 | 3.08 | 1.13 | 1.18 | −4 |
| Chile | Peso 1 670 | 3.07 | 519 | 544 | −5 |
| China | Yuán 11.0 | 1.41 | 3.42 | 7.77 | −56 |
| Colombia | Peso 6 900 | 3.06 | 2 143 | 2 254 | −5 |
| Corea del Sur | Won 2 900 | 3.08 | 901 | 942 | −4 |
| Costa Rica | Colón 1 130 | 2.18 | 351 | 519 | −32 |
| Dinamarca | Corona danesa 27.75 | 4.84 | 8.62 | 5.74 | +50 |
| Egipto | Libra 9.09 | 1.60 | 2.82 | 5.70 | −50 |
| Emiratos Árabes Unidos | Dirham 10.0 | 2.72 | 3.11 | 3.67 | −15 |
| Eslovaquia | Corona 57.98 | 2.14 | 18.0 | 27.2 | −34 |
| Estados Unidos | $3.22 | 3.22 | | | |
| Estonia | Corona 30 | 2.49 | 9.32 | 12.0 | −23 |
| Eurozona§ | €2.94 | 3.82 | 1.10** | 1.30** | +19 |
| Filipinas | Peso 85.0 | 1.74 | 26.4 | 48.9 | −46 |
| Gran Bretaña | £1.99 | 3.90 | 1.62‡ | 1.96‡ | +21 |
| Hong Kong | HK$12.0 | 1.54 | 3.73 | 7.81 | −52 |
| Hungría | Florín 590 | 3.00 | 183 | 197 | −7 |
| Indonesia | Rupia 15 900 | 1.75 | 4 938 | 9 100 | −46 |
| Islandia | Corona islándica 509 | 7.44 | 158 | 68.4 | +131 |
| Japón | ¥280 | 2.31 | 87.0 | 121 | −28 |
| Letonia | Lats 1.35 | 2.52 | 0.42 | 0.54 | −22 |
| Lituania | Litas 6.50 | 2.45 | 2.02 | 2.66 | −24 |
| Malasia | Ringgit 5.50 | 1.57 | 1.71 | 3.50 | −51 |
| México | Peso 29.0 | 2.66 | 9.01 | 10.9 | −17 |
| Noruega | Corona noruega 41.5 | 6.63 | 12.9 | 6.26 | +106 |
| Nueva Zelanda | NZ$4.60 | 3.16 | 1.43 | 1.45 | −2 |
| Pakistán | Rupia 140 | 2.31 | 43.5 | 60.7 | −28 |
| Paraguay | Guaraní 10 000 | 1.90 | 3 106 | 5 250 | −41 |
| Perú | Nuevo Sol 9.50 | 2.97 | 2.95 | 3.20 | −8 |
| Polonia | Zloty 6.90 | 2.29 | 2.14 | 3.01 | −29 |
| República Checa | Corona 52.1 | 2.41 | 16.2 | 21.6 | −25 |
| Rusia | Rublo 49.0 | 1.85 | 15.2 | 26.5 | −43 |
| Singapur | S$3.60 | 2.34 | 1.12 | 1.54 | −27 |
| Sri Lanka | Rupia 190 | 1.75 | 59.0 | 109 | −46 |
| Sudáfrica | Rand 15.5 | 2.14 | 4.81 | 7.25 | −34 |
| Suecia | Corona sueca 32.0 | 4.59 | 9.94 | 6.97 | −43 |
| Suiza | Franco suizo 6.30 | 5.05 | 1.96 | 1.25 | +57 |
| Tailandia | Baht 62.0 | 1.78 | 19.3 | 34.7 | −45 |
| Taiwan | Nuevo dólar taiwanés 75.0 | 2.28 | 23.3 | 32.9 | −29 |
| Turquía | Lira 4.55 | 3.22 | 1.41 | 1.41 | nil |
| Ucrania | Grivnia 9.00 | 1.71 | 2.80 | 5.27 | −47 |
| Uruguay | Peso 55.0 | 2.17 | 17.1 | 25.3 | −33 |
| Venezuela | Bolívar 6 800 | 1.58 | 2 112 | 4 307 | −51 |

* Paridad del poder adquisitivo: precio local entre el precio en Estados Unidos.
** Dólares por euro.
† Promedio de Nueva York, Chicago, San Francisco y Atlanta.
‡‡ Dólares por libra.
§ Promedio ponderado de precios en la eurozona.

*b*) ¿Los resultados de la regresión apoyan sus expectativas? ¿Qué prueba formal aplica para demostrar su hipótesis?

*c*) ¿*The Economist* debe continuar publicando el Índice Big Mac? ¿Por qué?

5.17. Consulte los datos del examen SAT del ejercicio 2.16. Suponga que desea predecir las calificaciones de los hombres en matemáticas ($Y$) con base en las calificaciones obtenidas por las mujeres en esa misma materia ($X$) con la siguiente regresión:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

*a*) Estime el modelo anterior.

*b*) De los residuos estimados, verifique si se mantiene el supuesto de normalidad.

*c*) Ahora pruebe la hipótesis de que $\beta_2 = 1$, es decir, que existe una correspondencia uno a uno entre las calificaciones de matemáticas obtenidas por los hombres y las mujeres.

*d*) Prepare la tabla ANOVA para este problema.

5.18. Repita el ejercicio del problema anterior pero ahora que $Y$ y $X$ representen las calificaciones obtenidas por hombres y mujeres en lectura crítica, respectivamente.

5.19. La tabla 5.10 presenta los datos anuales correspondientes al índice de precios al consumidor (IPC) y al índice de precios al mayoreo (IPM), también conocido como índice de precios al productor (IPP), de la economía estadounidense durante el periodo 1980-2006.

**TABLA 5.10**
**IPC e IPP, Estados Unidos, 1980-2006**

Fuente: *Economic Report of the President,* 2007, tablas B-62 y B-65.

| | IPC total | IPP (total de bienes terminados) |
|---|---|---|
| 1980 | 82.4 | 88.0 |
| 1981 | 90.9 | 96.1 |
| 1982 | 96.5 | 100.0 |
| 1983 | 99.6 | 101.6 |
| 1984 | 103.9 | 103.7 |
| 1985 | 107.6 | 104.7 |
| 1986 | 109.6 | 103.2 |
| 1987 | 113.6 | 105.4 |
| 1988 | 118.3 | 108.0 |
| 1989 | 124.0 | 113.6 |
| 1990 | 130.7 | 119.2 |
| 1991 | 136.2 | 121.7 |
| 1992 | 140.3 | 123.2 |
| 1993 | 144.5 | 124.7 |
| 1994 | 148.2 | 125.5 |
| 1995 | 152.4 | 127.9 |
| 1996 | 156.9 | 131.3 |
| 1997 | 160.5 | 131.8 |
| 1998 | 163.0 | 130.7 |
| 1999 | 166.6 | 133.0 |
| 2000 | 172.2 | 138.0 |
| 2001 | 177.1 | 140.7 |
| 2002 | 179.9 | 138.9 |
| 2003 | 184.0 | 143.3 |
| 2004 | 188.9 | 148.5 |
| 2005 | 195.3 | 155.7 |
| 2006 | 201.6 | 160.3 |

*a*) Grafique el IPC en el eje vertical y el IPM en el horizontal. *A priori*, ¿qué tipo de relación espera entre los dos índices? ¿Por qué?

*b*) Suponga que desea predecir un índice con base en el otro. ¿Cuál utilizaría como regresada y cuál como regresora? ¿Por qué?

*c*) Efectúe la regresión que eligió en el inciso *b*). Muestre los resultados usuales. Pruebe la hipótesis de que existe una relación uno a uno entre ambos índices.

*d*) De los residuos obtenidos de la regresión de *c*), ¿se puede abrigar la hipótesis de que el verdadero término de error está normalmente distribuido? Muestre las pruebas que utilizó.

5.20 La tabla 5.11 presenta datos sobre el índice de mortalidad por cáncer pulmonar (100 = promedio) y el índice de tabaquismo (100 = promedio) correspondientes a 25 grupos ocupacionales.

*a*) Trace la gráfica del índice de mortalidad por cáncer sobre el índice de tabaquismo. ¿Qué pauta general observa?

*b*) Sea $Y =$ índice de mortalidad por cáncer y $X =$ índice de tabaquismo; estime un modelo de regresión lineal y obtenga las estadísticas de regresión acostumbradas.

*c*) Pruebe la hipótesis que el tabaquismo no tiene influencia sobre el cáncer pulmonar con $\alpha = 5\%$.

*d*) ¿Cuáles son las ocupaciones de mayor riesgo en función de la mortalidad por cáncer pulmonar? ¿Puede ofrecer algunas razones de por qué ocurre así?

*e*) ¿Existe algún modo de incorporar la categoría de ocupación explícitamente en el análisis de regresión?

**TABLA 5.11**
**Tabaquismo y cáncer pulmonar**

Fuente: http://lib.stat.cmu.edu/DASL/Datafiles/SmokingandCancer.html.

| Ocupación | Tabaquismo | Cáncer |
|---|---|---|
| Agricultores, silvicultores, pescadores | 77 | 84 |
| Mineros y canteros | 137 | 116 |
| Productores de gas, coque y sustancias químicas | 117 | 123 |
| Productores de vidrio y cerámica | 94 | 128 |
| Trabajadores de hornos, forjas y fundiciones | 116 | 155 |
| Trabajadores de la industria eléctrica y electrónica | 102 | 101 |
| Ingeniería y oficios relacionados | 111 | 118 |
| Trabajadores de la industria maderera | 93 | 113 |
| Trabajadores de la industria del cuero | 88 | 104 |
| Trabajadores textiles | 102 | 88 |
| Trabajadores de la industria del vestido | 91 | 104 |
| Trabajadores de la industria de alimentos, bebidas y tabaco | 104 | 129 |
| Trabajadores de la industria de papel e impresos | 107 | 86 |
| Fabricantes de otros productos | 112 | 96 |
| Trabajadores de la industria de construcción | 113 | 144 |
| Pintores y decoradores | 110 | 139 |
| Conductores de vehículos de motor, grúas, etcétera | 125 | 113 |
| Jornaleros no incluidos en otras clasificaciones | 113 | 146 |
| Trabajadores de la industria de transporte y comunicaciones | 115 | 128 |
| Almacenistas, encargados de tiendas, etcétera | 105 | 115 |
| Trabajadores administrativos | 87 | 79 |
| Trabajadores de ventas | 91 | 85 |
| Trabajadores de servicios, deportes y recreación | 100 | 120 |
| Administradores y gerentes | 76 | 60 |
| Artistas y trabajadores profesionales y técnicos | 66 | 51 |

# Apéndice 5A

## 5A.1 Distribuciones de probabilidad relacionadas con la distribución normal

Las **distribuciones *t*, ji cuadrada** ($\chi^2$) y ***F***, cuyas características primordiales se analizan en el **apéndice A**, están estrechamente relacionadas con la distribución normal. En vista de que se hará un uso extensivo de estas distribuciones de probabilidad en los siguientes capítulos, se resumirán sus relaciones con la distribución normal en el siguiente teorema; las pruebas, que rebasan el objetivo de este libro, se pueden encontrar en la bibliografía.[1]

**Teorema 5.1.** Si $Z_1, Z_2, \ldots, Z_n$ son variables aleatorias distribuidas normalmente y de manera independiente de modo que $Z_i \sim N(\mu_i, \sigma_i^2)$, entonces la suma $Z = \sum k_i Z_i$, donde $k_i$ son constantes no todas nulas, está también normalmente distribuida, con una media $\sum k_i \mu_i$ y una varianza $\sum k_i^2 \sigma_i^2$; es decir, $Z \sim N(\sum k_i \mu_i, \sum k_i^2 \sigma_i^2)$. *Nota:* $\mu$ representa el valor medio.

En resumen, las combinaciones lineales de las variables normales están ellas mismas distribuidas normalmente. Por ejemplo, si $Z_1$ y $Z_2$ están distribuidas de manera normal e independiente conforme $Z_i \sim N(10, 2)$ y $Z_2 \sim N(8, 8.5)$, entonces la combinación lineal $Z = 0.8Z_1 + 0.2Z_2$ también está normalmente distribuida con una media $= 0.8(10) + 0.2(8) = 9.6$ y una varianza $= 0.64(2) + 0.04(1.5) = 1.34$; es decir, $Z \sim (9.6, 1.34)$.

**Teorema 5.2.** Si $Z_1, Z_2, \ldots, Z_n$ están normalmente distribuidas pero no son independientes, la suma $Z = \sum k_i Z_i$, donde las $k_i$ son constantes y no todas igual a cero, también está normalmente distribuida, con una media $\sum k_i \mu_i$ y una varianza $[\sum k_i^2 \sigma_i^2 + 2 \sum k_i k_j \operatorname{cov}(Z_i, Z_j), i \neq j]$.

Por tanto, si $Z_i \sim N(6, 2)$ y $Z_2 \sim N(7, 3)$ y $\operatorname{cov}(Z_1 Z_2) = 0.8$, entonces la combinación lineal $0.6Z_1 + 0.4Z_2$ también está normalmente distribuida con media $= 0.6(6) + 0.4(7) = 6.4$ y varianza $= [0.36(2) + 0.16(3) + 2(0.6)(0.4)(0.8)] = 1.584$.

**Teorema 5.3.** Si $Z_1, Z_2, \ldots, Z_n$ son variables aleatorias distribuidas de manera normal e independiente de manera que cada $Z_I \sim N(0, 1)$, es decir, es una variable normal estandarizada, entonces $\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ sigue la distribución ji cuadrada con $n$ gl. De manera simbólica, $\sum Z_i^2 \sim \chi_n^2$, donde $n$ representa los grados de libertad, gl.

En resumen, "la suma de los cuadrados de las variables normales estandarizadas independientes tiene una distribución ji cuadrada, con grados de libertad iguales al número de términos de la suma".[2]

**Teorema 5.4.** Si $Z_1, Z_2, \ldots, Z_n$ son variables aleatorias distribuidas independientemente y cada una sigue una distribución ji cuadrada con $k_i$ gl, entonces la suma $\sum Z_i = Z_1 + Z_2 + \cdots + Z_n$ también sigue una distribución ji cuadrada, con $k = \sum k_i$ gl.

Por tanto, si $Z_1$ y $Z_2$ son variables $\chi^2$ independientes con gl de $k_1$ y $k_2$, respectivamente, entonces $Z = Z_1 + Z_2$ también es una variable $\chi^2$ con $(k_1 + k_2)$ grados de libertad. Lo anterior se conoce como **propiedad reproductiva** de la distribución $\chi^2$.

[1] Hay pruebas de los distintos teoremas en Alexander M. Mood, Franklin A. Graybill y Duane C. Bose, *Introduction to the Theory of Statistics*, 3a. ed., McGraw-Hill, Nueva York, 1974, pp. 239-249.

[2] *Ibid.*, p. 243.

**Teorema 5.5.** Si $Z_1$ es una variable normal estandarizada [$Z_1 \sim N(0, 1)$] y otra variable $Z_2$ sigue la distribución ji cuadrada con $k$ gl y es independiente de $Z_1$, entonces la variable definida como

$$t = \frac{Z_1}{\sqrt{Z_2}/\sqrt{k}} = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}} = \frac{\text{Variable normal estandarizada}}{\sqrt{\text{Variable independiente ji cuadrada/gl}}} \sim t_k$$

sigue la distribución $t$ de Student, con $k$ gl. *Nota:* Esta distribución se analiza en el **apéndice A** y se ilustra en el capítulo 5.

A propósito, observe que, conforme $k$, los gl, se incrementa de manera indefinida (es decir, conforme $k \to \infty$), la distribución $t$ de Student se aproxima a la distribución normal estandarizada.[3] Por convención, la notación $t_k$ significa la distribución $t$ de Student o la variable con $k$ gl.

**Teorema 5.6.** Si $Z_1$ y $Z_2$ son variables ji cuadrada independientemente distribuidas, con $k_1$ y $k_2$ gl, respectivamente, entonces la variable

$$F = \frac{Z_1/k_1}{Z_2/k_2} \sim F_{k_1,k_2}$$

tiene la distribución $F$ con $k_1$ y $k_2$ grados de libertad, donde $k_1$ se conocen como los **grados de libertad del numerador** y $k_2$ como los **grados de libertad del denominador**.

De nuevo, por convención, la notación $F_{k_1,k_2}$ significa que es una variable $F$ con $k_1$ y $k_2$ grados de libertad, y los gl del numerador se colocan primero.

En otras palabras, el teorema 5.6 enuncia que la variable $F$ es sólo la razón de dos variables ji cuadrada distribuidas independientemente, divididas entre sus respectivos grados de libertad.

**Teorema 5.7.** El cuadrado de la variable $t$ (de Student), con $k$ gl, tiene una distribución $F$, con $k_1 = 1$ gl en el numerador y $k_2 = k$ gl en el denominador.[4] Es decir,

$$F_{1,k} = t_k^2$$

Observe que para que sea válida esta igualdad, los gl del numerador de la variable $F$ deben ser 1. Por tanto, $F_{1,4} = t_4^2$ o $F_{1,23} = t_{23}^2$, y así sucesivamente.

Como se mencionó, veremos la utilidad práctica del teorema anterior conforme avancemos.

**Teorema 5.8.** Para un valor grande de gl del denominador, la multiplicación de los gl del numerador por el valor $F$ es más o menos igual al valor de ji cuadrada con los gl del numerador. Así,

$$m\ F_{m,n} = \chi_m^2 \qquad \text{conforme } n \to \infty$$

**Teorema 5.9.** Para un número lo bastante grande de gl, la distribución ji cuadrada se aproxima mediante la distribución normal estándar de la siguiente forma:

$$Z = \sqrt{2\chi^2} - \sqrt{2k-1} \sim N(0, 1)$$

donde $k$ denota los gl.

[3] Para ver una prueba, consulte Henri Theil, *Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1978, pp. 237-245.

[4] Para ver una prueba, consulte las ecuaciones (5.3.2) y (5.9.1).

## 5A.2 Derivación de la ecuación (5.3.2)

Sea

$$Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\text{ee}\,(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{x_i^2}}{\sigma} \tag{1}$$

y

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{2}$$

Siempre que se conozca $\sigma$, $Z_1$ sigue una distribución normal estándar; es decir, $Z_1 \sim N(0, 1)$. (¿Por qué?) $Z_2$ sigue la distribución $\chi^2$ con $(n - 2)$ gl.[5] Además, puede demostrarse que $Z_2$ está independientemente distribuida de $Z_1$.[6] Por consiguiente, en virtud del teorema 5.5, la variable

$$t = \frac{Z_1\sqrt{n-2}}{\sqrt{Z_2}} \tag{3}$$

sigue la distribución $t$ con $n - 2$ gl. De la sustitución de (1) y (2) en (3), se obtiene la ecuación (5.3.2).

## 5A.3 Derivación de la ecuación (5.9.1)

La ecuación (1) muestra que $Z_1 \sim N(0, 1)$. Por consiguiente, por el teorema 5.3, la cantidad precedente

$$Z_1^2 = \frac{(\hat{\beta}_2 - \beta_2)^2 \sum x_i^2}{\sigma^2}$$

sigue la distribución $\chi^2$ con 1 gl. Como se mostró en la sección 5A.1,

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum \hat{u}_i^2}{\sigma^2}$$

también sigue la distribución $\chi^2$ con $n - 2$ gl. Además, como se vio en la sección 4.3, $Z_2$ está distribuida independientemente de $Z_1$. Entonces, al aplicar el teorema 5.6, se tiene que

$$F = \frac{Z_1^2/1}{Z_2/(n-2)} = \frac{(\hat{\beta}_2 - \beta_2)^2\left(\sum x_i^2\right)}{\sum \hat{u}_i^2/(n-2)}$$

sigue la distribución $F$ con 1 y $n - 2$ gl, respectivamente. Según la hipótesis nula $H_0$: $\beta_2 = 0$, la razón $F$ anterior se reduce a la ecuación (5.9.1).

## 5A.4 Derivación de las ecuaciones (5.10.2) y (5.10.6)

### Varianza de la predicción media

Dado $X_i = X_0$, la verdadera predicción media $E(Y_0 \mid X_0)$ está dada por

$$E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \tag{1}$$

[5] Para ver una prueba, consulte Robert V. Hogg y Allen T. Craig, *Introduction to Mathematical Statistics*, 2a. ed., Macmillan, Nueva York, 1965, p. 144.

[6] Para una demostración, véase J. Johnston, *Econometric Methods*, McGraw-Hill, 3a. ed., Nueva York, 1984, pp. 181-182. (Se requieren conocimientos de álgebra matricial para entender la prueba.)

Estimamos (1) a partir de

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{2}$$

Al obtener el valor esperado de (2), dado $X_0$, obtenemos

$$\begin{aligned} E(\hat{Y}_0) &= E(\hat{\beta}_1) + E(\hat{\beta}_2)X_0 \\ &= \beta_1 + \beta_2 X_0 \end{aligned}$$

porque $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores insesgados. Por consiguiente,

$$E(\hat{Y}_0) = E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \tag{3}$$

Es decir, $\hat{Y}_0$ es un predictor insesgado de $E(Y_0 \mid X_0)$.

Ahora, con la propiedad de que $\text{var}(a + b) = \text{var}(a) + \text{var}(b) + 2\,\text{cov}(a, b)$, obtenemos

$$\text{var}(\hat{Y}_0) = \text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2)X_0^2 + 2\,\text{cov}(\hat{\beta}_1\hat{\beta}_2)X_0 \tag{4}$$

Con las fórmulas para varianzas y covarianza de $\hat{\beta}_1$ y $\hat{\beta}_2$ dadas en (3.3.1), (3.3.3) y (3.3.9), y al reordenar términos, obtenemos

$$\text{var}(\hat{Y}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \qquad = (5.10.2)$$

## Varianza de la predicción individual

Deseamos predecir una $Y$ individual correspondiente a $X = X_0$, es decir, queremos obtener:

$$Y_0 = \beta_1 + \beta_2 X_0 + u_0 \tag{5}$$

Predecimos de la siguiente forma:

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{6}$$

El error de predicción, $Y_0 - \hat{Y}_0$, es

$$\begin{aligned} Y_0 - \hat{Y}_0 &= \beta_1 + \beta_2 X_0 + u_0 - (\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_0 + u_0 \end{aligned} \tag{7}$$

Por consiguiente,

$$\begin{aligned} E(Y_0 - \hat{Y}_0) &= E(\beta_1 - \hat{\beta}_1) + E(\beta_2 - \hat{\beta}_2)X_0 - E(u_0) \\ &= 0 \end{aligned}$$

porque $\hat{\beta}_1$, $\hat{\beta}_2$ son insesgados, $X_0$ es un número fijo y $E(u_0)$ es cero por suposición.

Al elevar (7) al cuadrado en ambos lados y tomar valores esperados, obtenemos $\text{var}(Y_0 - \hat{Y}_0) = \text{var}(\hat{\beta}_1) + X_0^2\,\text{var}(\hat{\beta}_2) + 2X_0\,\text{cov}(\beta_1, \beta_2) + \text{var}(u_0)$. Con las fórmulas de varianza y covarianza para $\hat{\beta}_1$ y $\hat{\beta}_2$ dadas antes, y al observar que var $(u_0 = \sigma^2)$, obtenemos

$$\text{var}(Y_0 - \hat{Y}_0) = \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \qquad = (5.10.6)$$

Capítulo 6

# Extensiones del modelo de regresión lineal con dos variables

Algunos aspectos del análisis de regresión lineal se insertan bien en el marco del modelo de regresión lineal con dos variables que hemos analizado hasta ahora. Primero consideraremos la **regresión a través del origen**, es decir, una situación en la cual el término del intercepto, $\beta_1$, está ausente del modelo. Luego veremos el tema de las **unidades de medición**, o la forma como se midieron $X$ y $Y$, y cómo un cambio en las unidades de medición afecta los resultados de la regresión. Por último, abordaremos el tema de la **forma funcional** del modelo de regresión lineal. Hasta el momento, consideramos modelos lineales en los parámetros y en las variables. Sin embargo, recuerde que la teoría de regresión de los capítulos anteriores sólo exige linealidad en los parámetros; las variables pueden o no entrar linealmente en el modelo. Al considerar modelos que son lineales en los parámetros pero no necesariamente en las variables, en este capítulo mostraremos la forma como el modelo de dos variables resuelve algunos problemas prácticos de interés.

Una vez entendidas las ideas de este capítulo, su extensión a los modelos de regresión múltiple es muy sencilla, como comprobaremos en los capítulos 7 y 8.

## 6.1 Regresión a través del origen

Hay ocasiones en las cuales la función de regresión poblacional (FRP) de dos variables adquiere la siguiente forma:

$$Y_i = \beta_2 X_i + u_i \tag{6.1.1}$$

En este modelo, el término del intercepto está ausente o es cero, lo cual explica el nombre: **regresión a través del origen**.

A manera de ilustración consideremos el modelo de asignación de precios de activos de capital (CAPM, del inglés *capital asset pricing model*) de la teoría moderna de portafolios, la cual, en su versión de prima por riesgo, se expresa como[1]

$$(\mathrm{ER}_i - r_f) = \beta_i(\mathrm{ER}_m - r_f) \tag{6.1.2}$$

[1] Véase Haim Levy y Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, Nueva Jersey, 1984, capítulo 14.

donde $ER_i$ = tasa esperada de rendimiento del título *i*.
$ER_m$ = tasa esperada de rendimiento del portafolios del mercado como la representa, por ejemplo, el índice compuesto de acciones S&P 500.
$r_f$ = tasa de rendimiento libre de riesgo, por ejemplo, el rendimiento de los bonos del Tesoro estadounidense a 90 días.
$\beta_i$ = el coeficiente Beta, una medida de riesgo sistemático, es decir, el riesgo que no se ha eliminado con la diversificación. Asimismo, es una medida del grado en el cual la *i*-ésima tasa de rendimiento del título se mueve con el mercado. Un $\beta_i > 1$ implica un título volátil o riesgoso, mientras que $\beta_i < 1$ es un título seguro. (*Nota:* No confunda esta $\beta_i$ con el coeficiente de la pendiente de la regresión con dos variables, $\beta_2$.)

Si los mercados de capitales funcionan de manera eficiente, el CAPM postula que la prima esperada por el riesgo del título ($= ER_i - r_f$) es igual a ese coeficiente $\beta$ del título multiplicado por la prima esperada del riesgo del mercado ($= ER_m - r_f$). Si el CAPM se mantiene se da la situación de la figura 6.1. La línea que aparece en la figura se conoce como **línea del mercado de valores** (LMV).

Para fines empíricos, (6.1.2) suele expresarse así:

$$R_i - r_f = \beta_i(R_m - r_f) + u_i \qquad \textbf{(6.1.3)}$$

o

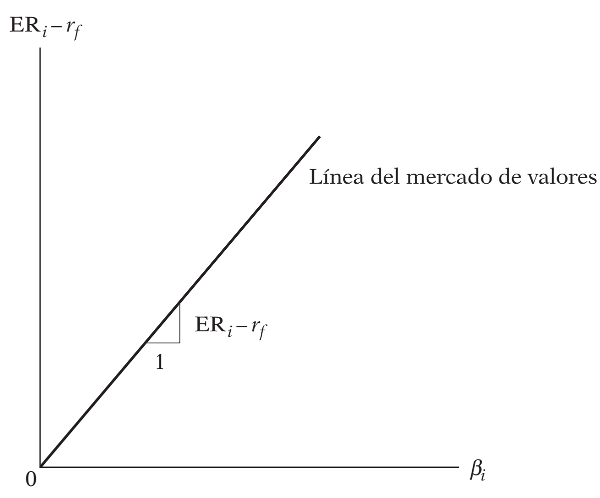
$$R_i - r_f = \alpha_i + \beta_i(R_m - r_f) + u_i \qquad \textbf{(6.1.4)}$$

Este último modelo se conoce como el **Modelo del Mercado**.[2] Si el CAPM es válido, se espera que $\alpha_i$ sea cero. (Véase la figura 6.2.)

Observe que en (6.1.4) la variable dependiente, *Y*, es $(R_i - r_f)$, y la variable explicativa, *X*, es $\beta_i$, el coeficiente de volatilidad, y *no* $(R_m - r_f)$. Por consiguiente, para realizar la regresión (6.1.4), se debe estimar primero $\beta_i$, el cual se obtiene por lo general de la **línea característica**, como describimos en el ejercicio 5.5. (Para mayores detalles véase el ejercicio 8.28.)

Como muestra este ejemplo, algunas veces la teoría que sirve de base requiere que el término del intercepto esté ausente del modelo. La hipótesis del ingreso permanente de Milton Friedman, que afirma que el consumo permanente es proporcional al ingreso permanente, es otro caso en el que el modelo de intercepto cero puede ser apropiado, como también en la teoría del análisis

**FIGURA 6.1**
Riesgo sistemático.




[2] Véase, por ejemplo, Diana R. Harrington, *Modern Portfolio Theory and the Capital Asset Pricing Model: A User's Guide*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1983, p. 71.

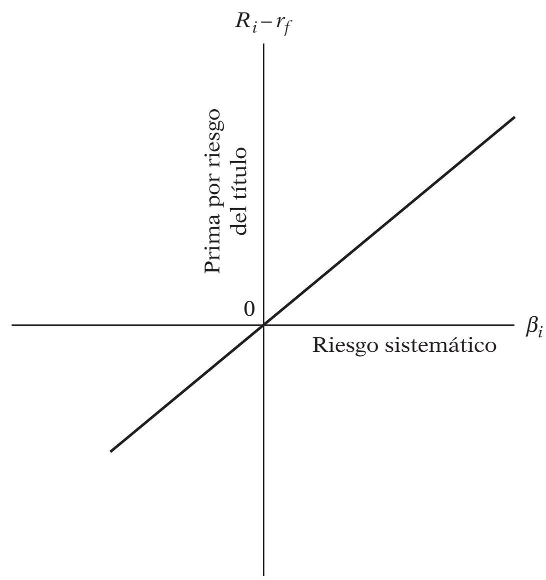



**FIGURA 6.2**
El Modelo del Mercado de la teoría de portafolios (con el supuesto de que $\alpha_i = 0$).

de costos, que postula que la variable costo de producción es proporcional a la producción; y algunas versiones de la teoría monetarista que afirman que la tasa de cambio de los precios (es decir, la tasa de inflación) es proporcional a la tasa de cambio de la oferta monetaria.

¿Cómo se estiman modelos como (6.1.1) y qué problemas presentan? Para responder, primero escribimos la FRM de (6.1.1), a saber:

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i \tag{6.1.5}$$

Ahora aplicamos el método MCO a (6.1.5) y se obtienen las siguientes fórmulas para $\hat{\beta}_2$ y su varianza (las pruebas se presentan en el apéndice 6A, sección 6A.1):

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{6.1.6}$$

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2} \tag{6.1.7}$$

donde $\sigma^2$ se estima con

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1} \tag{6.1.8}$$

Es interesante comparar estas fórmulas con las obtenidas cuando se incluye el término del intercepto en el modelo:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{3.1.6}$$

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

Deben ser obvias las diferencias entre estos dos conjuntos de fórmulas: en el modelo sin término de intercepto se utilizan sumas de cuadrados **simples** y productos cruzados, pero en el modelo con intercepto, se utilizan sumas de cuadrados ajustadas (de la media) y productos cruzados. Segundo, los gl para calcular $\hat{\sigma}^2$ son $(n - 1)$ en el primer caso y $(n - 2)$ en el segundo. (¿Por qué?)

Aunque el modelo sin intercepto o con intercepto cero puede ser apropiado en algunas ocasiones, deben observarse algunas características de este modelo. Primero, $\sum \hat{u}_i$, que es siempre cero en el modelo con intercepto (el modelo convencional), no necesita serlo cuando ese término está ausente. En resumen, $\sum \hat{u}_i$ no necesita ser cero en la regresión a través del origen. Segundo, $r^2$, el coeficiente de determinación presentado en el capítulo 3, que siempre es no negativo en el modelo convencional, en ocasiones puede volverse *negativo* en el modelo sin intercepto. Este resultado anómalo surge porque el $r^2$ que presentamos en el capítulo 3 supone explícitamente que el intercepto está incluido en el modelo. Por consiguiente, el $r^2$ calculado convencionalmente puede no ser apropiado en los modelos de regresión a través del origen.[3]

## $r^2$ para el modelo de regresión a través del origen

Como recién mencionamos y más adelante analizaremos en mayor detalle en el apéndice 6A, sección 6A.1, el $r^2$ convencional del capítulo 3 no es apropiado en regresiones que no incluyan o no consideren el intercepto. Pero se puede calcular para tales modelos, lo que se conoce como el $\boldsymbol{r^2}$ **simple**, el cual se define como

$$r^2\,\text{simple} = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2} \tag{6.1.9}$$

*Nota:* Se trata de sumas de cuadrados simples (es decir, no corregidas por la media) y de productos cruzados.

A pesar de que este $r^2$ simple satisface la relación $0 < r^2 < 1$, no es directamente comparable con el valor $r^2$ convencional. Por esta razón, algunos autores no presentan el valor $r^2$ en los modelos de regresión con intercepto cero.

Debido a las características especiales de este modelo, se debe tener mucho cuidado al utilizar el modelo de regresión con intercepto cero. *A menos que haya una expectativa a priori muy sólida*, es aconsejable apegarse al modelo convencional con presencia de intercepto. Esto tiene una doble ventaja. Primero, si se incluye en el modelo el término del intercepto pero es estadísticamente no significativo (es decir, estadísticamente igual a cero), para todos los fines prácticos se tiene una regresión a través del origen.[4] Segundo y más importante, si el modelo sí tiene un intercepto pero insistimos en ajustar una regresión a través del origen, cometeríamos un **error de especificación**. Veremos esto en detalle en el capítulo 7.

**EJEMPLO 6.1**

La tabla 6.1 presenta datos mensuales sobre los rendimientos excedentes $Y_t$(%) de un índice de 104 acciones del sector de bienes de consumo cíclico y los rendimientos excedentes $X_t$(%) del índice de todo el mercado de valores en el Reino Unido, correspondientes al periodo 1980-1999, para un total de 240 observaciones.[5] Por *rendimientos excedentes* se entiende el rendimiento superior al que ofrece un activo sin riesgo (véase el modelo CAPM).

[3] Para un análisis adicional, véase Dennis J. Aigner, *Basic Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1971, pp. 85-88.

[4] Henri Theil señala que si el intercepto efectivamente está ausente, el coeficiente de la pendiente puede estimarse con mucha más precisión que cuando el término del intercepto está incluido. Véase su *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, p. 76. Véase también el ejemplo numérico a continuación.

[5] Estos datos, provenientes originalmente del banco de datos *DataStream,* se reproducen de Christiaan Heij *et al., Econometric Methods with Applications in Business and Economics,* Oxford University Press, Oxford, Reino Unido, 2004.

**TABLA 6.1**

| OBS | *Y* | *X* | OBS | *Y* | *X* |
|---|---|---|---|---|---|
| 1980:01 | 6.08022852 | 7.263448404 | 1984:12 | 3.52786616 | 3.191554763 |
| 1980:02 | −0.924185461 | 6.339895504 | 1985:01 | 4.554587707 | 3.907838688 |
| 1980:03 | −3.286174252 | −9.285216834 | 1985:02 | 5.365478677 | −1.708567484 |
| 1980:04 | 5.211976571 | 0.793290771 | 1985:03 | 4.525231564 | 0.435218492 |
| 1980:05 | −16.16421111 | −2.902420985 | 1985:04 | 2.944654344 | 0.958067845 |
| 1980:06 | −1.054703649 | 8.613150875 | 1985:05 | −0.268599528 | 1.095477375 |
| 1980:07 | 11.17237699 | 3.982062848 | 1985:06 | −3.661040481 | −6.816108909 |
| 1980:08 | −11.06327551 | −1.150170907 | 1985:07 | −4.540505062 | 2.785054354 |
| 1980:09 | −16.77699609 | 3.486125868 | 1985:08 | 9.195292816 | 3.900209023 |
| 1980:10 | −7.021834032 | 4.329850278 | 1985:09 | −1.894817019 | −4.203004414 |
| 1980:11 | −9.71684668 | 0.936875279 | 1985:10 | 12.00661274 | 5.60179802 |
| 1980:12 | 5.215705717 | −5.202455846 | 1985:11 | 1.233987382 | 1.570093976 |
| 1981:01 | −6.612000956 | −2.082757509 | 1985:12 | −1.446329607 | −1.084427121 |
| 1981:02 | 4.264498443 | 2.728522893 | 1986:01 | 6.023618851 | 0.778669473 |
| 1981:03 | 4.916710821 | 0.653397106 | 1986:02 | 10.51235756 | 6.470651262 |
| 1981:04 | 22.20495946 | 6.436071962 | 1986:03 | 13.40071024 | 8.953781192 |
| 1981:05 | −11.29868524 | −4.259197932 | 1986:04 | −7.796262998 | −2.387761685 |
| 1981:06 | −5.770507783 | 0.543909707 | 1986:05 | 0.211540446 | −2.873838588 |
| 1981:07 | −5.217764717 | −0.486845933 | 1986:06 | 6.471111064 | 3.440269098 |
| 1981:08 | 16.19620175 | 2.843999508 | 1986:07 | −9.037475168 | −5.891053375 |
| 1981:09 | −17.16995395 | −16.4572142 | 1986:08 | −5.47838091 | 6.375582004 |
| 1981:10 | 1.105334728 | 4.468938171 | 1986:09 | −6.756881852 | −5.734839396 |
| 1981:11 | 11.6853367 | 5.885519658 | 1986:10 | −2.564960223 | 3.63088408 |
| 1981:12 | −2.301451728 | −0.390698164 | 1986:11 | 2.456599468 | −1.31606687 |
| 1982:01 | 8.643728679 | 2.499567896 | 1986:12 | 1.476421303 | 3.521601216 |
| 1982:02 | −11.12907503 | −4.033607075 | 1987:01 | 17.0694004 | 8.673412896 |
| 1982:03 | 1.724627956 | 3.042525777 | 1987:02 | 7.565726727 | 6.914361923 |
| 1982:04 | 0.157879967 | 0.734564665 | 1987:03 | −3.239325817 | −0.460660854 |
| 1982:05 | −1.875202616 | 2.779732288 | 1987:04 | 3.662578335 | 4.295976077 |
| 1982:06 | −10.62481767 | −5.900116576 | 1987:05 | 7.157455113 | 7.719692529 |
| 1982:07 | −5.761135416 | 3.005344385 | 1987:06 | 4.774901623 | 3.039887622 |
| 1982:08 | 5.481432596 | 3.954990619 | 1987:07 | 4.23770166 | 2.510223804 |
| 1982:09 | −17.02207459 | 2.547127067 | 1987:08 | −0.881352219 | −3.039443563 |
| 1982:10 | 7.625420708 | 4.329008106 | 1987:09 | 11.49688416 | 3.787092018 |
| 1982:11 | −6.575721646 | 0.191940594 | 1987:10 | −35.56617624 | −27.86969311 |
| 1982:12 | −2.372829861 | −0.92167555 | 1987:11 | −14.59137369 | −9.956367094 |
| 1983:01 | 17.52374936 | 3.394682577 | 1987:12 | 14.87271664 | 7.975865948 |
| 1983:02 | 1.354655809 | 0.758714353 | 1988:01 | 1.748599294 | 3.936938398 |
| 1983:03 | 16.26861049 | 1.862073664 | 1988:02 | −0.606016446 | −0.32797064 |
| 1983:04 | −6.074547158 | 6.797751341 | 1988:03 | −6.078095523 | −2.161544202 |
| 1983:05 | −0.826650702 | −1.699253628 | 1988:04 | 3.976153828 | 2.721787842 |
| 1983:06 | 3.807881996 | 4.092592402 | 1988:05 | −1.050910058 | −0.514825422 |
| 1983:07 | 0.57570091 | −2.926299262 | 1988:06 | 3.317856956 | 3.128796482 |
| 1983:08 | 3.755563441 | 1.773424306 | 1988:07 | 0.407100105 | 0.181502075 |
| 1983:09 | −5.365927271 | −2.800815667 | 1988:08 | −11.87932524 | −7.892363786 |
| 1983:10 | −3.750302815 | −1.505394995 | 1988:09 | −8.801026046 | 3.347081899 |
| 1983:11 | 4.898751703 | 4.18696284 | 1988:10 | 6.784211277 | 3.158592144 |
| 1983:12 | 4.379256151 | 1.201416981 | 1988:11 | −10.20578119 | −4.816470363 |
| 1984:01 | 16.56016188 | 6.769320788 | 1988:12 | −6.73805381 | −0.008549997 |
| 1984:02 | 1.523127464 | −1.686027417 | 1989:01 | 12.83903643 | 13.46098219 |
| 1984:03 | 1.0206078 | 5.245806105 | 1989:02 | 3.302860922 | −0.764474692 |
| 1984:04 | −3.899307684 | 1.728710264 | 1989:03 | −0.155918301 | 2.298491097 |
| 1984:05 | −14.32501615 | −7.279075595 | 1989:04 | 3.623090767 | 0.762074588 |
| 1984:06 | 3.056627177 | −0.77947067 | 1989:05 | −1.167680873 | −0.495796117 |
| 1984:07 | −0.02153592 | −2.439634487 | 1989:06 | −1.221603303 | 1.206636013 |
| 1984:08 | 3.355102212 | 8.445977813 | 1989:07 | 5.262902744 | 4.637026116 |
| 1984:09 | 0.100006778 | 1.221080129 | 1989:08 | 4.845013219 | 2.680874116 |
| 1984:10 | 1.691250318 | 2.733386772 | 1989:09 | −5.069564838 | −5.303858035 |
| 1984:11 | 8.20075301 | 5.12753329 | 1989:10 | −13.57963526 | −7.210655599 |

(*continúa*)

**TABLA 6.1**
(*continuación*)

| OBS | Y | X | OBS | Y | X |
|---|---|---|---|---|---|
| 1989:11 | 1.100607603 | 5.350185944 | 1994:12 | −4.225370964 | 0.264280259 |
| 1989:12 | 4.925083189 | 4.106245855 | 1995:01 | −6.302392617 | −2.420388431 |
| 1990:01 | −2.532068851 | −3.629547374 | 1995:02 | 1.27867637 | 0.138795213 |
| 1990:02 | −6.601872876 | −5.205804299 | 1995:03 | 10.90890516 | 3.231656585 |
| 1990:03 | −1.023768943 | −2.183244863 | 1995:04 | 2.497849434 | 2.215804682 |
| 1990:04 | −7.097917266 | −5.408563794 | 1995:05 | 2.891526594 | 3.856813589 |
| 1990:05 | 6.376626925 | 10.57599169 | 1995:06 | −3.773000069 | −0.952204306 |
| 1990:06 | 1.861974711 | −0.338612099 | 1995:07 | 8.776288715 | 4.020036363 |
| 1990:07 | −5.591527585 | −2.21316202 | 1995:08 | 2.88256097 | 1.423600345 |
| 1990:08 | −15.31758975 | −8.476177427 | 1995:09 | 2.14691333 | −0.037912571 |
| 1990:09 | −10.17227358 | −7.45941471 | 1995:10 | −4.590104662 | −1.17655329 |
| 1990:10 | −2.217396045 | −0.085887763 | 1995:11 | −1.293255187 | 3.760277356 |
| 1990:11 | 5.974205798 | 5.034770534 | 1995:12 | −4.244101531 | 0.434626357 |
| 1990:12 | −0.857289036 | −1.767714908 | 1996:01 | 6.647088904 | 1.906345103 |
| 1991:01 | −3.780184589 | 0.189108456 | 1996:02 | 1.635900742 | 0.301898961 |
| 1991:02 | 20.64721437 | 10.38741504 | 1996:03 | 7.8581899 | −0.314132324 |
| 1991:03 | 10.94068018 | 2.921913827 | 1996:04 | 0.789544896 | 3.034331741 |
| 1991:04 | −3.145639589 | 0.971720188 | 1996:05 | −0.907725397 | −1.497346299 |
| 1991:05 | −3.142887645 | −0.4317819 | 1996:06 | −0.392246948 | −0.894676854 |
| 1991:06 | −1.960866141 | −3.342924986 | 1996:07 | −1.035896351 | −0.532816274 |
| 1991:07 | 7.330964031 | 5.242811509 | 1996:08 | 2.556816005 | 3.863737088 |
| 1991:08 | 7.854387926 | 2.880654691 | 1996:09 | 3.131830038 | 2.118254897 |
| 1991:09 | 2.539177843 | −1.121472224 | 1996:10 | −0.020947358 | −0.853553262 |
| 1991:10 | −1.233244642 | −3.969577956 | 1996:11 | −5.312287782 | 1.770340939 |
| 1991:11 | −11.7460404 | −5.707995062 | 1996:12 | −5.196176326 | 1.702551635 |
| 1991:12 | 1.078226286 | 1.502567049 | 1997:01 | −0.753247124 | 3.465753348 |
| 1992:01 | 5.937904622 | 2.599565094 | 1997:02 | −2.474343938 | 1.115253221 |
| 1992:02 | 4.113184542 | 0.135881087 | 1997:03 | 2.47647802 | −2.057818461 |
| 1992:03 | −0.655199392 | −6.146138064 | 1997:04 | −1.119104196 | 3.57089955 |
| 1992:04 | 15.28430278 | 10.45736831 | 1997:05 | 3.352076269 | 1.953480438 |
| 1992:05 | 3.994517585 | 1.415987046 | 1997:06 | −1.910172239 | 2.458700404 |
| 1992:06 | −11.94450998 | −8.261109424 | 1997:07 | 0.142814607 | 2.992341297 |
| 1992:07 | −2.530701327 | −3.778812167 | 1997:08 | 10.50199263 | −0.457968038 |
| 1992:08 | −9.842366221 | −5.386818488 | 1997:09 | 12.98501943 | 8.111278967 |
| 1992:09 | 18.11573724 | 11.19436372 | 1997:10 | −4.134761655 | −6.967124504 |
| 1992:10 | 0.200950206 | 3.999870038 | 1997:11 | −4.148579856 | −0.155924791 |
| 1992:11 | 1.125853097 | 3.620674752 | 1997:12 | −1.752478236 | 3.853283433 |
| 1992:12 | 7.639180786 | 2.887222251 | 1998:01 | −3.349121498 | 7.379466014 |
| 1993:01 | 2.919569408 | 1.336746091 | 1998:02 | 14.07471304 | 4.299097886 |
| 1993:02 | −1.062404105 | 1.240273846 | 1998:03 | 7.791650968 | 3.410780517 |
| 1993:03 | 1.292641409 | 0.407144312 | 1998:04 | 5.154679109 | −0.081494993 |
| 1993:04 | 0.420241384 | −1.734930047 | 1998:05 | 3.293686179 | −1.613131159 |
| 1993:05 | −2.514080553 | 1.111533687 | 1998:06 | −13.25461802 | −0.397288954 |
| 1993:06 | 0.419362276 | 1.354127742 | 1998:07 | −7.714205916 | −2.237365283 |
| 1993:07 | 4.374024535 | 1.943061568 | 1998:08 | −15.26340483 | −12.4631993 |
| 1993:08 | 1.733528075 | 4.961979827 | 1998:09 | −15.22865141 | −5.170734985 |
| 1993:09 | −3.659808969 | −1.618729936 | 1998:10 | 15.96218038 | 11.70544788 |
| 1993:10 | 5.85690764 | 4.215408608 | 1998:11 | −8.684089113 | −0.380200223 |
| 1993:11 | −1.365550294 | 1.880360165 | 1998:12 | 17.13842369 | 4.986705187 |
| 1993:12 | −1.346979017 | 5.826352413 | 1999:01 | −1.468448611 | 2.493727994 |
| 1994:01 | 12.89578758 | 2.973540693 | 1999:02 | 8.5036 | 0.937105259 |
| 1994:02 | −5.346700561 | −5.479858563 | 1999:03 | 10.8943073 | 4.280082506 |
| 1994:03 | −7.614726564 | −5.784547088 | 1999:04 | 13.03497394 | 3.960824402 |
| 1994:04 | 10.22042923 | 1.157083438 | 1999:05 | −5.654671597 | −4.499198079 |
| 1994:05 | −6.928422261 | −6.356199493 | 1999:06 | 8.321969316 | 3.656745699 |
| 1994:06 | −5.065919037 | −0.843583888 | 1999:07 | 0.507652273 | −2.503971473 |
| 1994:07 | 7.483498556 | 5.779953224 | 1999:08 | −5.022980561 | −0.121901923 |
| 1994:08 | 1.828762662 | 3.298130184 | 1999:09 | −2.305448839 | −5.388032432 |
| 1994:09 | −5.69293279 | −7.110010085 | 1999:10 | −1.876879466 | 4.010989716 |
| 1994:10 | −2.426962489 | 2.968005597 | 1999:11 | 1.348824769 | 6.265312975 |
| 1994:11 | 2.125100668 | −1.531245158 | 1999:12 | −2.64164938 | 4.045658427 |

**EJEMPLO 6.1**
(*continuación*)

En primer lugar ajustamos el modelo (6.1.3) a estos datos. Con *EViews6* obtuvimos los siguientes resultados de regresión, que se presentan en el formato estándar de *EViews.*

Variable dependiente: *Y*
Método: mínimos cuadrados
Muestra: 1980M01 1999M12
Observaciones incluidas: 240

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| ***X*** | **1.155512** | **0.074396** | **15.53200** | **0.0000** |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.500309 | Media de la variable dependiente | 0.499826 |
| *R* cuadrada ajustada† | 0.500309 | Desviación estándar de la variable dependiente | 7.849594 |
| Error estándar de regresión | 5.548786 | Estadístico de Durbin-Watson* | 1.972853 |
| Suma de cuadrados de residuos | 7 358.578 | | |

*Estudiaremos este estadístico en el capítulo 12.
† Véase el capítulo 7.

Como muestran estos resultados, el coeficiente de la pendiente (el coeficiente Beta) es muy significativo, pues su valor *p* es muy pequeño. La interpretación en este caso es que si la tasa excedente del mercado aumenta un punto porcentual, el rendimiento excedente del índice del sector de bienes de consumo aumenta alrededor de 1.15 puntos porcentuales. El coeficiente de la pendiente no es sólo estadísticamente significativo, sino que es significativamente mayor que 1 (¿puede verificar esto?). Si un coeficiente Beta es mayor que 1, se dice que ese título (en este caso, un portafolios de 104 acciones) es volátil; se mueve más que proporcionalmente con el índice general del mercado de valores. Sin embargo, este resultado no debe sorprender, porque en este ejemplo se consideran acciones del sector de bienes de consumo cíclico, como los bienes duraderos de uso doméstico, automóviles, textiles y equipo deportivo.

Si ajustamos el modelo (6.1.4), obtenemos los siguientes resultados:

Variable dependiente: *Y*
Método: mínimos cuadrados
Muestra: 1980M01 1999M12
Observaciones incluidas: 240

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| ***C*** | **–0.447481** | **0.362943** | **–1.232924** | **0.2188** |
| ***X*** | **1.171128** | **0.075386** | **15.53500** | **0.0000** |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.503480 | Media de la variable dependiente | 0.499826 |
| *R* cuadrada ajustada | 0.501394 | Desviación estándar de la variable dependiente | 7.849594 |
| Error estándar de regresión | 5.542759 | Estadístico de Durbin-Watson | 1.984746 |
| Suma de cuadrados de residuos | 7 311.877 | Probabilidad (estadístico *F*) | 0.000000 |
| Estadístico F | 241.3363 | | |

En estos resultados observamos que el intercepto no es estadísticamente diferente de cero, aunque el coeficiente de la pendiente (el coeficiente Beta) es muy significativo estadísticamente. Esto indica que el modelo de regresión a través del origen se ajusta bien a los datos. Además, en términos estadísticos, no hay diferencia entre los valores del coeficiente de la pendiente en los dos modelos. Observe que el error estándar del coeficiente de la pendiente en el modelo de regresión a través del origen es un poco menor que el del modelo con el intercepto presente, lo cual apoya el argumento de Theil de la nota 4. Aun en este caso, el coeficiente de la pendiente es estadísticamente mayor que 1, lo que una vez más confirma que los rendimientos de las acciones del sector de bienes de consumo cíclico son volátiles.

A propósito, observe que el valor de $r^2$ para el modelo de regresión a través del origen debe tomarse con ciertas reservas, pues la fórmula tradicional de $r^2$ no es aplicable en tales modelos. Sin embargo, *EViews* presenta de manera habitual el valor estándar de $r^2$, incluso para estos modelos.

## 6.2 Escalas y unidades de medición

Para entender las ideas de esta sección, considere la información de la tabla 6.2, referente a la inversión doméstica privada bruta (IDPB) de Estados Unidos y al producto interno bruto (PIB) en miles de millones y en millones de dólares de 2000 ajustados por la inflación.

Suponga que en la regresión de la IDPB sobre el PIB, un investigador utiliza información medida en miles de millones de dólares y otro expresa estos datos en millones de dólares. ¿Serán iguales los resultados de la regresión en ambos casos? De no ser así, ¿qué resultados deben usarse? En resumen, ¿las unidades con que se mide la variable regresada y la(s) variable(s) regresora(s) influyen de algún modo en los resultados de la regresión? De ser así, ¿qué curso razonable debe seguirse en la selección de las unidades de medición para el análisis de regresión? Para responder estas preguntas, procedamos sistemáticamente. Sea

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{6.2.1}$$

donde $Y$ = IDPB y $X$ = PIB. Defina

$$Y_i^* = w_1 Y_i \tag{6.2.2}$$

$$X_i^* = w_2 X_i \tag{6.2.3}$$

donde $w_1$ y $w_2$ son constantes, denominadas **factores de escala**; $w_1$ puede ser igual o diferente a $w_2$.

De (6.2.2) y (6.2.3) es claro que $Y_i^*$ y $X_i^*$ son $Y_i$ y $X_i$ *reescaladas*. Por tanto, si $Y_i$ y $X_i$ se miden en miles de millones de dólares y se desea expresarlas en millones de dólares, se tendrá $Y_i^* =$ 1 000 $Y_i$ y $X_i^* =$ 1 000 $X_i$; aquí $w_1 = w_2 =$ 1 000.

Ahora considere la regresión con las variables $Y_i^*$ y $X_i^*$:

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{6.2.4}$$

donde $Y_i^* = w_1 Y_i$, $X_i^* = w_2 X_i$ y $\hat{u}_i^* = w_1 \hat{u}_i$. (¿Por qué?)

**TABLA 6.2**
**Inversión nacional privada bruta y PIB, Estados Unidos, 1990-2005 (miles de millones de dólares [de 2000] ajustados por la inflación, salvo donde se indica lo contrario; datos trimestrales con tasas anuales ajustadas por estacionalidad)**

Fuente: *Economic Report of the President,* 2007, tabla B-2, p. 328.

| Año | $IDPB_{mm}$ | $IDPB_m$ | $PIB_{mm}$ | $PIB_m$ |
|---|---|---|---|---|
| 1990 | 886.6 | 886 600.0 | 7 112.5 | 7 112 500.0 |
| 1991 | 829.1 | 829 100.0 | 7 100.5 | 7 100 500.0 |
| 1992 | 878.3 | 878 300.0 | 7 336.6 | 7 336 600.0 |
| 1993 | 953.5 | 953 500.0 | 7 532.7 | 7 532 700.0 |
| 1994 | 1 042.3 | 1 042 300.0 | 7 835.5 | 7 835 500.0 |
| 1995 | 1 109.6 | 1 109 600.0 | 8 031.7 | 8 031 700.0 |
| 1996 | 1 209.2 | 1 209 200.0 | 8 328.9 | 8 328 900.0 |
| 1997 | 1 320.6 | 1 320 600.0 | 8 703.5 | 8 703 500.0 |
| 1998 | 1 455.0 | 1 455 000.0 | 9 066.9 | 9 066 900.0 |
| 1999 | 1 576.3 | 1 576 300.0 | 9 470.3 | 9 470 300.0 |
| 2000 | 1 679.0 | 1 679 000.0 | 9 817.0 | 9 817 000.0 |
| 2001 | 1 629.4 | 1 629 400.0 | 9 890.7 | 9 890 700.0 |
| 2002 | 1 544.6 | 1 544 600.0 | 10 048.8 | 10 048 800.0 |
| 2003 | 1 596.9 | 1 596 900.0 | 10 301.0 | 10 301 000.0 |
| 2004 | 1 713.9 | 1 713 900.0 | 10 703.5 | 10 703 500.0 |
| 2005 | 1 842.0 | 1 842 000.0 | 11 048.6 | 11 048 600.0 |

*Nota:* $IDPB_{mm}$ = inversión doméstica privada bruta (miles de millones de dólares de 2000).
$IDPB_m$ = inversiones nacionales privadas brutas (millones de dólares de 2000).
$PIB_{mm}$ = producto interno bruto (miles de millones de dólares de 2000).
$PIB_m$ = producto interno bruto (millones de dólares de 2000).

Deseamos encontrar las relaciones entre los siguientes pares:

1. $\hat{\beta}_1$ y $\hat{\beta}_1^*$
2. $\hat{\beta}_2$ y $\hat{\beta}_2^*$
3. var $(\hat{\beta}_1)$ y var$(\hat{\beta}_1^*)$
4. var $(\hat{\beta}_2)$ y var$(\hat{\beta}_2^*)$
5. $\hat{\sigma}^2$ y $\hat{\sigma}^{*2}$
6. $r_{xy}^2$ y $r_{x^*y^*}^2$

De la teoría de mínimos cuadrados, sabemos (véase el capítulo 3) que

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} \tag{6.2.5}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{6.2.6}$$

$$\text{var}\,(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2} \cdot \sigma^2 \tag{6.2.7}$$

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{6.2.8}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{6.2.9}$$

Del mismo modo, al aplicar el método MCO a (6.2.4), obtenemos

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^*\bar{X}^* \tag{6.2.10}$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \tag{6.2.11}$$

$$\text{var}\,(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n\sum x_i^{*2}} \cdot \sigma^{*2} \tag{6.2.12}$$

$$\text{var}\,(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \tag{6.2.13}$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \tag{6.2.14}$$

Con estos resultados es fácil establecer relaciones entre estos dos conjuntos de parámetros estimados. Todo lo que se debe hacer es recordar las siguientes relaciones: $Y_i^* = w_1 Y_i$ (o $y_i^* = w_1 y_i$); $X_i^* = w_2 X_i$ (o $x_i^* = w_2 x_i$); $\hat{u}_i^* = w_1 \hat{u}_i$; $\bar{Y}^* = w_1 \bar{Y}$; y $\bar{X}^* = w_2 \bar{X}$. Con estas definiciones, el lector puede verificar fácilmente que

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2 \tag{6.2.15}$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \tag{6.2.16}$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \tag{6.2.17}$$

$$\text{var}\,(\hat{\beta}_1^*) = w_1^2\, \text{var}\,(\hat{\beta}_1) \tag{6.2.18}$$

$$\text{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \text{var}(\hat{\beta}_2) \qquad (6.2.19)$$

$$r_{xy}^2 = r_{x^*y^*}^2 \qquad (6.2.20)$$

De los resultados anteriores debe quedar claro que, con los resultados de regresión basados en una escala de medición, se pueden obtener los resultados basados en otra, una vez que se conozcan los factores de escala, $w$. En la práctica, sin embargo, se deben escoger las unidades de medición en forma razonable; no tiene objeto manejar todos esos ceros al expresar números en millones o en miles de millones de dólares.

De los resultados de (6.2.15) hasta (6.2.20) se derivan fácilmente algunos casos especiales. Por ejemplo, si $w_1 = w_2$, es decir, si son idénticos los factores de escala, el coeficiente de la pendiente y su error estándar permanecen inalterados en el cambio de escala de $(Y_i, X_i)$ a $(Y_i^*, X_i^*)$, lo cual intuitivamente debería ser claro. Sin embargo, el intercepto y su error estándar están multiplicados por $w_1$. Si la escala $X$ no se cambia (es decir, $w_2 = 1$), pero la escala $Y$ se cambia por el factor $w_1$, el coeficiente de la pendiente, al igual que el intercepto y sus errores estándar respectivos, se multiplican por el mismo factor $w_1$. Por último, si la escala $Y$ permanece inalterada (es decir, $w_1 = 1$), pero la escala $X$ se cambia por el factor $w_2$, el coeficiente de la pendiente y su error estándar se multiplican por el factor $(1/w_2)$, pero el coeficiente del intercepto y su error estándar permanecen inalterados.

Sin embargo, debe observarse que la transformación de la escala $(Y, X)$ a la escala $(Y^*, X^*)$ no afecta las propiedades de los estimadores de MCO analizadas en los capítulos anteriores.

**EJEMPLO 6.2**
*Relación entre la IDPB y el PIB, Estados Unidos, 1990-2005*

Para demostrar los resultados teóricos anteriores, consideremos de nuevo los datos presentados en la tabla 6.2 y examinemos los siguientes resultados (las cifras entre paréntesis son los errores estándar estimados).

Si las escalas de la IDPB y del PIB están en miles de millones de dólares:

$$\begin{aligned}\widehat{\text{IDPB}}_t &= -926.090 + 0.2535\ \text{PIB}_t \\ \text{ee} &= \ (116.358) \quad (0.0129) \qquad r^2 = 0.9648\end{aligned} \qquad (6.2.21)$$

Si las escalas de la IDPB y del PIB están en millones de dólares:

$$\begin{aligned}\widehat{\text{IDPB}}_t &= -926\ 090 + 0.2535\ \text{PIB}_t \\ \text{ee} &= \ (116.358) \quad (0.0129) \qquad r^2 = 0.9648\end{aligned} \qquad (6.2.22)$$

Observe que el intercepto, lo mismo que su error estándar, es 1 000 veces los valores correspondientes de la regresión (6.2.21) (observe que $w_1 = 1\ 000$ al pasar de miles de millones a millones de dólares), pero el coeficiente de la pendiente, al igual que su error estándar, permanecen sin cambio, como lo afirma la teoría.

La IDPB en miles de millones de dólares y el PIB en millones de dólares:

$$\begin{aligned}\widehat{\text{IDPB}}_t &= -926.090 + 0.0002535\ \text{PIB}_t \\ \text{ee} &= \ (116.358) \quad (0.0000129) \qquad r^2 = 0.9648\end{aligned} \qquad (6.2.23)$$

Como se esperaba, el coeficiente de la pendiente, al igual que su error estándar, es (1/1 000) de su valor en (6.2.21), pues sólo se modificó la escala de $X$, es decir, del PIB.

La IDPB en millones de dólares y el PIB en miles de millones de dólares:

$$\begin{aligned}\widehat{\text{IDPB}}_t &= -926\ 090 + 253.524\ \text{PIB}_t \\ \text{ee} &= \ (116\ 358.7) \quad (12.9465) \qquad r^2 = 0.9648\end{aligned} \qquad (6.2.24)$$

De nuevo, observe que tanto el intercepto como el coeficiente de la pendiente y sus errores estándar respectivos son 1 000 veces sus valores en (6.2.21), lo cual concuerda con los resultados teóricos.

Note que, en todas las regresiones presentadas antes, el valor de $r^2$ permanece constante, lo cual no sorprende debido a que el valor $r^2$ es *invariable* respecto de los cambios en las unidades de medición, pues es un número puro o adimensional.

### Advertencia sobre la interpretación

Como el coeficiente de la pendiente, $\beta_2$, es tan sólo la tasa de cambio, ésta se mide en las unidades de la razón

$$\frac{\text{Unidades de la variable dependiente}}{\text{Unidades de la variable explicativa}}$$

Así, en la regresión (6.2.21), la interpretación del coeficiente de la pendiente 0.2535 es que si el PIB cambia en una unidad, de 1 000 millones de dólares, la IDPB cambia en promedio en 0.2535 miles de millones de dólares. En la regresión (6.2.23), una unidad de cambio en el PIB, que es 1 millón de dólares, induce en promedio a un cambio de 0.0002535 miles de millones de dólares en la IDPB. Los dos resultados son por supuesto idénticos en sus efectos del PIB sobre la IDPB, simplemente están expresados en diferentes unidades de medición.

## 6.3 Regresión sobre variables estandarizadas

En la sección anterior vimos que las unidades con que se expresan la variable independiente (regresora) y la dependiente (regresada) influyen en la interpretación de los coeficientes de regresión. Esto se evita si ambas variables (regresora y regresada) se expresan como *variables estandarizadas*. Se dice que una variable es estandarizada si se resta el valor de la media de esta variable de sus valores individuales y se divide esa diferencia entre la desviación estándar de la variable.

Así, en la regresión de $Y$ y $X$, si las redefinimos como:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad \textbf{(6.3.1)}$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \quad \textbf{(6.3.2)}$$

donde $\bar{Y}$ = media muestral de $Y$, $S_Y$ = desviación estándar muestral de $Y$, $\bar{X}$ = media muestral de $X$ y $S_X$ = desviación estándar muestral de $X$; las variables $Y_i^*$ y $X_i^*$ se llaman **variables estandarizadas**.

*Una propiedad interesante de una variable estandarizada es que el valor de su media siempre es cero y que su desviación estándar siempre es 1*. (Para comprobar lo anterior, véase el apéndice 6A, sección 6A.2.)

Como resultado, no importa en qué unidades se expresen ambas variables (la regresada y la regresora). En consecuencia, en lugar de llevar a cabo la regresión estándar (bivariada):

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \textbf{(6.3.3)}$$

podemos realizar la regresión sobre las variables estandarizadas de la siguiente manera:

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* \quad \textbf{(6.3.4)}$$

$$= \beta_2^* X_i^* + u_i^* \quad \textbf{(6.3.5)}$$

pues resulta sencillo mostrar que, en la regresión que involucra a la regresada estandarizada y a la(s) regresora(s) estandarizada(s), el término del intercepto siempre es cero.[6] Los coeficientes de regresión de las variables estandarizadas, denotados por $\beta_1^*$ y $\beta_2^*$, se conocen en la bibliografía como los **coeficientes beta**.[7] Por cierto, observe que (6.3.5) es una regresión a través del origen.

¿Cómo se interpretan los coeficientes beta? La interpretación es que si la regresora (estandarizada) se incrementa una desviación estándar, en promedio, la regresada (estandarizada) aumenta $\beta_2^*$ unidades de desviación estándar. Por tanto, a diferencia del modelo tradicional (6.3.3), se mide el efecto no en términos de las unidades originales en las expresadas $X$ y $Y$, sino en unidades de desviación estándar.

Para mostrar la diferencia entre (6.3.3) y (6.3.5) regresaremos al ejemplo de la IDPB y el PIB de la sección anterior. Los resultados de (6.2.21), ya examinados, se reproducen a continuación:

$$\begin{aligned} \widehat{\text{IDPB}}_t &= -926.090 \;+\; 0.2535\ \text{PIB}_t \\ \text{ee} &= \quad (116.358) \qquad (0.0129) \qquad r^2 = 0.9648 \end{aligned} \tag{6.3.6}$$

donde la IDPB y el PIB se miden en miles de millones de dólares.

Los resultados que corresponden a (6.3.5) son los siguientes, en donde las variables con asterisco son variables estandarizadas:

$$\begin{aligned} \widehat{\text{IDPB}}_t^* &= \;0.9822\ \text{PIB}_t^* \\ \text{ee} &= (0.0485) \end{aligned} \tag{6.3.7}$$

Ya sabemos interpretar (6.3.6): si el PIB se incrementa un dólar, la IDPB aumenta, en promedio, 30 centavos. ¿Y qué pasa con (6.3.7)? Aquí se interpreta como sigue: si el PIB (estandarizado) se incrementara una desviación estándar, en promedio, la IDPB (estandarizada) aumentaría casi 0.94 desviaciones estándar.

¿Cuál es la ventaja del modelo de regresión estandarizado respecto del modelo tradicional? Ésta se manifiesta mejor cuando hay más de una regresora, tema que analizaremos en el capítulo 7. Al estandarizar todas las regresoras, quedan expresadas en una misma base y por consiguiente se pueden comparar de manera directa. Si el coeficiente de una regresora estandarizada es mayor que el de otra regresora estandarizada que aparece en ese modelo, esta última contribuye relativamente más a la explicación de la regresada de lo que contribuye la primera. En otras palabras, los coeficientes beta sirven como medida de la fuerza relativa de las diversas regresoras. Profundizaremos más en este tema en los dos siguientes capítulos.

Antes de dar por terminado este asunto, vale la pena un par de observaciones. Primero, para la regresión estandarizada (6.3.7), no se dio el valor $r^2$ porque es una regresión a través del origen, para la cual no se aplica la $r^2$ usual, como se señaló en la sección 6.1. Segundo, existe una relación interesante entre los coeficientes $\beta$ del modelo convencional y los coeficientes beta. Para el caso bivariado, la relación es como sigue:

$$\hat{\beta}_2^* = \hat{\beta}_2 \left( \frac{S_x}{S_y} \right) \tag{6.3.8}$$

donde $S_x$ = la desviación estándar muestral de la regresora $X$ y $S_y$ = la desviación estándar muestral de la regresada. Por consiguiente, se pueden intercambiar los $\beta$ con los coeficientes beta si se conoce la desviación estándar (muestral) de la regresora y de la regresada. En el siguiente capítulo veremos que esta relación se cumple también para la regresión múltiple. Se deja como ejercicio para el lector verificar la ecuación (6.3.8) para este ejemplo ilustrativo.

[6] Recuerde, de la ecuación (3.1.7), que el intercepto es igual al valor de la media de la variable dependiente menos la pendiente multiplicada por el valor de la media de la regresora. No obstante, para las variables estandarizadas, los valores de la media de la variable dependiente y de la regresora son cero. Por tanto, el valor del intercepto también es cero.

[7] No confunda estos coeficientes beta con los coeficientes beta de la teoría financiera.

## 6.4 Formas funcionales de los modelos de regresión

Como mencionamos en el capítulo 2, este texto trata sobre todo con modelos lineales en los parámetros, que pueden ser o no lineales en las variables. En las secciones que siguen consideraremos algunos modelos de regresión muy comunes, que pueden ser no lineales en las variables pero sí lineales en los parámetros, o que pueden serlo mediante transformaciones apropiadas de las variables. En particular, analizaremos los siguientes modelos de regresión:

1. El modelo log-lineal.
2. Modelos semilogarítmicos.
3. Modelos recíprocos.
4. El modelo logarítmico recíproco.

Ahora analizaremos las características especiales de cada modelo, los casos en los cuales su uso es apropiado y la forma de estimarlos. Cada modelo se ilustra con ejemplos apropiados.

## 6.5 Cómo medir la elasticidad: modelo log-lineal

Considere el siguiente modelo, conocido como **modelo de regresión exponencial**:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.5.1}$$

que puede expresarse también como[8]

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \tag{6.5.2}$$

donde ln = logaritmo natural (es decir, logaritmo en base $e$ y donde $e = 2.718$).[9]

Si escribimos (6.5.2) como

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.5.3}$$

donde $\alpha = \ln \beta_1$, este modelo es lineal en los parámetros $\alpha$ y $\beta_2$, lineal en los logaritmos de las variables $Y$ y $X$, y se estima por regresión MCO. Debido a esta linealidad, tales modelos se denominan modelos **log-log**, **doble-log** o **log-lineales**. Véase el apéndice 6A.3, donde se explican las propiedades de los logaritmos.

Si se cumplen los supuestos del modelo clásico de regresión lineal, los parámetros de (6.5.3) se estiman por el método MCO, considerando que

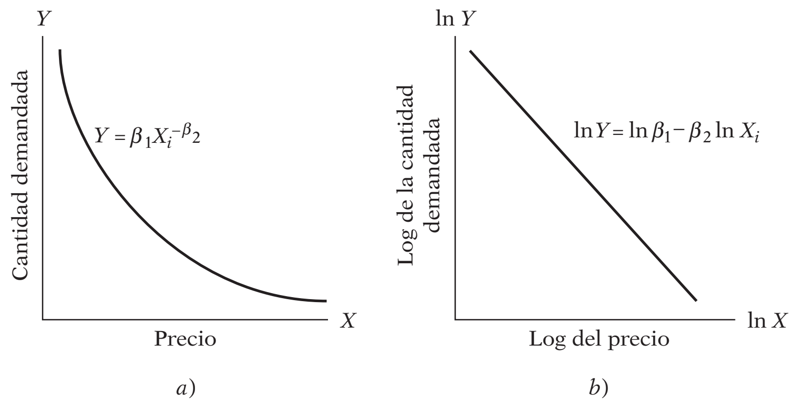
$$Y_i^* = \alpha + \beta_2 X_i^* + u_i \tag{6.5.4}$$

donde $Y_i^* = \ln Y_i$ y $X_i^* = \ln X_i$. Los estimadores de MCO obtenidos, $\hat{\alpha}$ y $\hat{\beta}_2$, serán los mejores estimadores lineales insesgados de $\alpha$ y $\beta_2$, respectivamente.

[8] Observe estas propiedades de los logaritmos: 1) $\ln(AB) = \ln A + \ln B$, 2) $\ln(A/B) = \ln A - \ln B$ y 3) $\ln(A^k) = k \ln A$, suponiendo que $A$ y $B$ son positivos, y donde $k$ es alguna constante.

[9] En la práctica se pueden utilizar logaritmos comunes, es decir, logaritmos base 10. La relación entre el logaritmo natural y el logaritmo común es: $\ln_e X = 2.3026 \log_{10} X$. Por convención, ln significa logaritmo natural y log significa logaritmo base 10; por tanto, no hay necesidad de escribir los subíndices $e$ y 10 explícitamente.

**FIGURA 6.3**
Modelo de elasticidad constante.




Una característica atractiva del modelo log-log, que lo ha hecho muy popular en el trabajo empírico, es que el coeficiente de la pendiente $\beta_2$ mide la **elasticidad** de $Y$ respecto de $X$, es decir, el cambio porcentual en $Y$ ante un pequeño cambio porcentual en $X$.[10] Así, si $Y$ representa la cantidad demandada de un bien y $X$ su precio unitario, $\beta_2$ mide la elasticidad-precio de la demanda, parámetro de gran interés en economía. Si la relación entre la cantidad demandada y el precio es como se muestra en la figura 6.3*a*, la transformación doble-log de la figura 6.3*b* dará entonces la estimación de la elasticidad-precio $(-\beta_2)$.

Pueden observarse dos características especiales del modelo log-lineal: el modelo supone que el coeficiente de la elasticidad entre $Y$ y $X$, $\beta_2$, permanece constante a través del tiempo (¿por qué?), de aquí su otro nombre, **modelo de elasticidad constante**.[11] En otras palabras, como lo indica la figura 6.3*b*, el cambio en ln $Y$ por unidad de cambio en ln $X$ (es decir, la elasticidad, $\beta_2$) permanece igual sin importar en cuál ln $X$ se mida la elasticidad. Otro aspecto del modelo es que, a pesar de que $\hat{\alpha}$ y $\hat{\beta}_2$ son estimadores insesgados de $\alpha$ y $\beta_2$, $\beta_1$ (el parámetro del modelo original) al estimarse como $\hat{\beta}_1 = \text{antilog}\ (\hat{\alpha})$ es, en sí, un estimador sesgado. En la mayor parte de los problemas prácticos, sin embargo, el término del intercepto es de importancia secundaria y no es necesario preocuparse por obtener este estimador insesgado.[12]

[10] El coeficiente de elasticidad, en la notación del cálculo, se define como $(dY/Y)/(dX/X) = [(dY/dX)(X/Y)]$. Los lectores familiarizados con el cálculo diferencial se darán cuenta fácilmente de que $\beta_2$ es, en efecto, el coeficiente de elasticidad.

*Nota técnica:* El lector con una buena formación en cálculo notará que $d(\ln X)/dX = 1/X$ o $d(\ln X) = dX/X$, es decir, para cambios infinitesimalmente pequeños (observe el operador diferencial $d$), un cambio en ln $X$ es igual al cambio relativo o proporcional en $X$. En la práctica, sin embargo, si el cambio en $X$ es pequeño, esta relación se escribe como: cambio en ln $X \doteq$ cambio relativo en $X$, donde $\doteq$ significa aproximadamente igual. Así, para cambios pequeños,

$$(\ln X_t - \ln X_{t-1}) \doteq (X_t - X_{t-1})/X_{t-1} = \text{cambio relativo en } X$$

A propósito, el lector debe observar estos términos, a los que se hace referencia frecuentemente: 1) **cambio absoluto**, 2) **cambio relativo** o **proporcional** y 3) **cambio porcentual** o **tasa de crecimiento porcentual**. Así, $(X_t - X_{t-1})$ representa un cambio absoluto, $(X_t - X_{t-1})/X_{t-1} = (X_t/X_{t-1} - 1)$ es un cambio relativo o proporcional, y $[(X_t - X_{t-1})/X_{t-1}]100$ es el cambio porcentual, o la tasa de crecimiento. $X_t$ y $X_{t-1}$ son los valores actual y anterior de la variable $X$, respectivamente.

[11] Un modelo de elasticidad constante permitirá obtener un cambio constante en el ingreso total ante un cambio porcentual dado en precios sin importar el nivel absoluto del precio. Los lectores deben contrastar este resultado con las condiciones de elasticidad que implica una función de demanda lineal simple, $Y_i = \beta_1 + \beta_2 X_i + u_i$. Sin embargo, una función lineal simple permite obtener un cambio constante en la cantidad generada por un cambio unitario en el precio. Confronte esto con lo que implica el modelo log-lineal para un cambio nominal en los precios.

[12] En relación con la naturaleza del sesgo y lo que puede hacerse al respecto, véase Arthur S. Goldberger, *Topics in Regression Analysis,* Macmillan, Nueva York, 1978, p. 120.

En el modelo de dos variables, la forma más simple de decidir si el modelo log-lineal se ajusta a los datos es graficar el diagrama de dispersión de ln $Y_i$ frente a ln $X_i$ y ver si las observaciones caen más o menos sobre una línea recta, como en la figura 6.3*b*.

*Advertencia:* El lector debe tener presente la distinción entre un cambio porcentual y uno en puntos porcentuales. Por ejemplo, la tasa de desempleo a menudo se expresa en forma de porcentaje; por decir, una tasa de desempleo de 6%. Si esta tasa aumenta a 8%, se dice que el cambio en puntos porcentuales de la tasa de desempleo es 2, mientras que el cambio porcentual de la tasa de desempleo es $(8-6)/6$, o alrededor de 33%. Por consiguiente, hay que tener cuidado cuando se trabaja con cambios porcentuales y cambios en puntos porcentuales, pues son dos conceptos muy diferentes.

**EJEMPLO 6.3**

*Gasto en bienes duraderos en relación con el gasto de consumo personal total*

La tabla 6.3 presenta datos sobre el gasto de consumo personal total (GCPERT), el gasto en bienes duraderos (GASBD), el gasto en bienes perecederos (GASBPER) y el gasto en servicios (GASERV), todos medidos en miles de millones de dólares de 2000.[13]

Suponga que deseamos calcular la elasticidad del gasto en bienes duraderos respecto del gasto de consumo personal total. Al graficar el logaritmo del gasto en bienes duraderos contra el logaritmo del gasto de consumo personal total, observará que la relación entre las dos variables es lineal. Por tanto, el modelo del doble logaritmo puede resultar adecuado. Los resultados de la regresión son:

$$\begin{aligned}\widehat{\ln \text{GASBD}_t} &= -7.5417 + 1.6266 \ln \text{GCPERT}_t \\ \text{ee} &= (0.7161) \quad (0.0800) \\ t &= (-10.5309)^* \quad (20.3152)^* \qquad r^2 = 0.9695\end{aligned} \tag{6.5.5}$$

donde * indica que el valor *p* es en extremo pequeño.

**TABLA 6.3**

**Gasto personal total y categorías (miles de millones de dólares de 2000 ajustados por la inflación; datos trimestrales con tasas anuales ajustadas por estacionalidad)**

Fuentes: Departamento de Comercio, Oficina de Análisis Económico, *Economic Report of the President*, 2007, tabla B-17, p. 347.

| Año o trimestre | GASERV | GASBD | GASBPER | GCPERT |
|---|---|---|---|---|
| 2003-I | 4 143.3 | 971.4 | 2 072.5 | 7 184.9 |
| 2003-II | 4 161.3 | 1 009.8 | 2 084.2 | 7 249.3 |
| 2003-III | 4 190.7 | 1 049.6 | 2 123.0 | 7 352.9 |
| 2003-IV | 4 220.2 | 1 051.4 | 2 132.5 | 7 394.3 |
| 2004-I | 4 268.2 | 1 067.0 | 2 155.3 | 7 479.8 |
| 2004-II | 4 308.4 | 1 071.4 | 2 164.3 | 7 534.4 |
| 2004-III | 4 341.5 | 1 093.9 | 2 184.0 | 7 607.1 |
| 2004-IV | 4 377.4 | 1 110.3 | 2 213.1 | 7 687.1 |
| 2005-I | 4 395.3 | 1 116.8 | 2 241.5 | 7 739.4 |
| 2005-II | 4 420.0 | 1 150.8 | 2 268.4 | 7 819.8 |
| 2005-III | 4 454.5 | 1 175.9 | 2 287.6 | 7 895.3 |
| 2005-IV | 4 476.7 | 1 137.9 | 2 309.6 | 7 910.2 |
| 2006-I | 4 494.5 | 1 190.5 | 2 342.8 | 8 003.8 |
| 2006-II | 4 535.4 | 1 190.3 | 2 351.1 | 8 055.0 |
| 2006-III | 4 566.6 | 1 208.8 | 2 360.1 | 8 111.2 |

*Nota:* Véase la tabla B-2, que contiene datos sobre el gasto de consumo personal total correspondientes a 1959-1989.
GASERV = gasto en servicios (miles de millones de dólares de 2000).
GASBD = gasto en bienes duraderos (miles de millones de dólares de 2000).
GASBPER = gasto en bienes perecederos (miles de millones de dólares de 2000).
GCPERT = gasto de consumo personal total (miles de millones de dólares de 2000).

(*continúa*)

[13] Los bienes duraderos son vehículos automotores y refacciones, muebles y equipo doméstico; los bienes perecederos son comida, ropa, gasolina, aceite, combustible de petróleo y carbón mineral; y los servicios son vivienda, electricidad y gas, transporte y atención médica.

**EJEMPLO 6.3** (*continuación*)

Como muestran estos resultados, la elasticidad de GASBD respecto de GCPERT es de casi 1.63, lo que indica que si el gasto personal total aumenta 1%, en promedio, el gasto en bienes duraderos se incrementará casi 1.63%. En consecuencia, el gasto en bienes duraderos es muy sensible a los cambios en el gasto de consumo personal. Por esta razón, los productores de bienes duraderos siguen muy de cerca los cambios en el ingreso personal y el gasto de consumo personal. En el ejercicio 6.18 se pide al lector que realice un ejercicio similar para el gasto en bienes perecederos.

## 6.6 Modelos semilogarítmicos: log-lin y lin-log

### Cómo medir la tasa de crecimiento: modelo log-lin

A los economistas, comerciantes y gobiernos con frecuencia les interesa encontrar la tasa de crecimiento de ciertas variables económicas, como población, PNB, oferta monetaria, empleo, productividad y déficit comercial.

Suponga que deseamos conocer la tasa de crecimiento del gasto de consumo personal en servicios para los datos de la tabla 6.3. Sea $Y_t$ el gasto real en servicios en el tiempo $t$ y $Y_0$ el valor inicial del gasto en servicios (es decir, el valor al final del cuarto trimestre de 2002). Recordará la muy conocida fórmula del interés compuesto, vista en los cursos básicos de economía.

$$Y_t = Y_0(1 + r)^t \tag{6.6.1}$$

donde $r$ es la tasa de crecimiento compuesta de $Y$ (es decir, a través del tiempo). Con el logaritmo natural de (6.6.1), escribimos

$$\ln Y_t = \ln Y_0 + t \ln (1 + r) \tag{6.6.2}$$

Ahora, con

$$\beta_1 = \ln Y_0 \tag{6.6.3}$$

$$\beta_2 = \ln (1 + r) \tag{6.6.4}$$

escribimos (6.6.2) así

$$\ln Y_t = \beta_1 + \beta_2 t \tag{6.6.5}$$

Al agregar el término de perturbación a (6.6.5), obtenemos[14]

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \tag{6.6.6}$$

Este modelo es como cualquier otro modelo de regresión lineal en el sentido de que los parámetros $\beta_1$ y $\beta_2$ son lineales. La única diferencia es que la variable dependiente o regresada es el logaritmo de $Y$ y la regresora o variable explicativa es el "tiempo", que adquiere valores de 1, 2, 3, etcétera.

Los modelos como (6.6.6) se denominan **modelos semilog** porque sólo una variable (en este caso, la regresada) aparece en forma logarítmica. Para fines descriptivos, un modelo en el cual la variable regresada es logarítmica se denomina **modelo log-lin**. Más adelante consideraremos un modelo en el cual la variable regresada es lineal pero la(s) regresora(s) es (son) logarítmica(s): un modelo **lin-log**.

[14] Agregamos el término de error porque la fórmula de interés compuesto no se cumple con exactitud. La razón de agregar el error después de la transformación logarítmica se expone en la sección 6.8.

Antes de presentar los resultados de la regresión, examinemos las propiedades del modelo (6.6.5). En este modelo, *el coeficiente de la pendiente mide el cambio proporcional constante o relativo en Y para un cambio absoluto dado en el valor de la regresora* (en este caso, la variable $t$), es decir,[15]

$$\beta_2 = \frac{\text{cambio relativo en regresada}}{\text{cambio absoluto en la regresora}} \qquad \textbf{(6.6.7)}$$

Si multiplicamos el cambio relativo en $Y$ por 100, (6.6.7) dará entonces el cambio porcentual, o la *tasa de crecimiento*, en $Y$ ocasionada por un cambio absoluto en $X$, la variable regresora. Es decir, 100 por $\beta_2$ da como resultado la tasa de crecimiento en $Y$; 100 por $\beta_2$ se conoce en la bibliografía como la **semielasticidad** de $Y$ respecto de $X$. (Pregunta: Para conocer la elasticidad, ¿qué debemos hacer?)[16]

**EJEMPLO 6.4**

*Tasa de crecimiento del gasto en servicios*

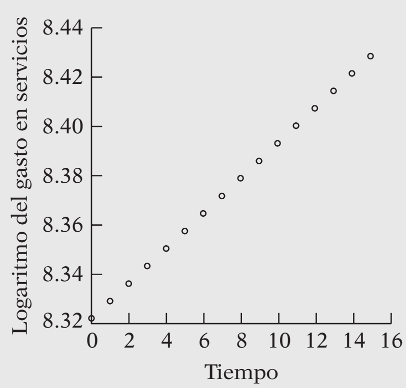
Para ilustrar el modelo de crecimiento (6.6.6), considere los datos sobre el gasto en servicios proporcionados en la tabla 6.3. Los resultados de la regresión a través del tiempo ($t$) son los siguientes:

$$\begin{aligned} \widehat{\ln \text{GES}_t} &= \quad 8.3226 \quad + \quad 0.00705t \\ \text{ee} &= \quad (0.0016) \qquad (0.00018) \qquad r^2 = 0.9919 \\ t &= (5\,201.625)^* \quad (39.1667)^* \end{aligned} \qquad \textbf{(6.6.8)}$$

*Nota:* GES significa gasto en servicios, y el asterisco (*) denota que el valor $p$ es en extremo pequeño.

La interpretación de la ecuación (6.6.8) es que durante los periodos trimestrales de 2003-I a 2006-III, el gasto en servicios se incrementó con una tasa (trimestral) de 0.705%. Aproximadamente, esto equivale a un crecimiento anual de 2.82%. Como 8.3226 = log de GES al comienzo del periodo de análisis, si se toma su antilogaritmo obtenemos 4 115.96 (miles de millones de dólares) como el valor inicial de GES (es decir, el valor a principios de 2003). En la figura 6.4 se ilustra la línea de regresión obtenida con la ecuación (6.6.8).

**FIGURA 6.4**




[15] Mediante cálculo diferencial se demuestra que $\beta_2 = d(\ln Y)/dX = (1/Y)(dY/dX) = (dY/Y)/dX$, que no es otra cosa que la ecuación (6.6.7). Para cambios pequeños en $Y$ y en $X$, esta relación puede aproximarse mediante

$$\frac{(Y_t - Y_{t-1})/Y_{t-1}}{(X_t - X_{t-1})}$$

*Nota:* Aquí, $X = t$.

[16] Véanse varias fórmulas de crecimiento en el apéndice 6A.4.

### *Tasas de crecimiento instantánea y compuesta*

El coeficiente de la variable de tendencia del modelo de crecimiento (6.6.6), $\beta_2$, da la tasa de crecimiento **instantánea** (en un momento dado) y no la **compuesta** (durante un periodo). Pero esta última se calcula fácilmente a partir de (6.6.4). Para ello, se obtiene el antilogaritmo de la $\beta_2$ estimada, se resta 1 y se multiplica la diferencia por 100. Por tanto, para el ejemplo ilustrativo, el coeficiente estimado de la pendiente es 0.00743. Así, [antilog(0.00746) − 1] = 0.00746 o 0.746%. En consecuencia, en el ejemplo ilustrativo, la *tasa compuesta de crecimiento* de gastos en servicios fue de casi 0.746% por trimestre, lo cual es un poco mayor que la tasa de crecimiento instantánea de 0.743%. Lo anterior se debe, por supuesto, al efecto de la composición.

### *Modelo de tendencia lineal*

En lugar de estimar el modelo (6.6.6), los investigadores algunas veces estiman el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 t + u_t \tag{6.6.9}$$

Es decir, en lugar de regresar el log de $Y$ sobre el tiempo, regresan $Y$ sobre el tiempo, donde $Y$ es la variable regresada en consideración. Un modelo de este tipo se denomina **modelo de tendencia lineal**, y la variable tiempo $t$ se conoce como *variable de tendencia*. Si el coeficiente de la pendiente en (6.6.9) es positivo, existe una **tendencia creciente** en $Y$, mientras que si es negativa, existe una **tendencia decreciente** en $Y$.

Para los datos sobre el gasto en servicios que analizamos antes, los resultados de ajustar el modelo de tendencia lineal (6.6.9) son los siguientes:

$$\begin{aligned}\widehat{\text{GES}}_t &= 4\,111.545 \;+\; 30.674t \\ t &= (655.5628) \quad (44.4671) \qquad r^2 = 0.9935\end{aligned} \tag{6.6.10}$$

En contraste con la ecuación (6.6.8), la interpretación de la ecuación (6.6.10) es la siguiente: durante los periodos trimestrales de 2003-I a 2006-III, en promedio, el gasto en servicios se incrementó con una tasa absoluta (*nota*: no relativa) de alrededor de 30 000 millones de dólares por trimestre. Es decir, hubo una tendencia creciente en el gasto en servicios.

La elección entre el modelo de crecimiento (6.6.8) y el modelo de tendencia lineal (6.6.10) dependerá de que el interés recaiga en el cambio relativo o absoluto del gasto en servicios, aunque, para propósitos de comparación, es el cambio relativo el que tiene mayor importancia. A propósito, *observe que no se pueden comparar los valores de $r^2$ de los modelos (6.6.8) y (6.6.10), pues las regresadas o variables dependientes son diferentes en los dos modelos*. Veremos en el capítulo 7 la forma de comparar las $r^2$ de los modelos como el (6.6.8) y el (6.6.10).

## El modelo lin-log

A diferencia del modelo de crecimiento recién estudiado, en el cual nos interesaba encontrar el crecimiento porcentual en $Y$ ante un cambio unitario absoluto en $X$, ahora deseamos encontrar el cambio absoluto en $Y$ debido a un cambio porcentual en $X$. Un modelo que cumple este propósito se escribe como

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{6.6.11}$$

Con fines descriptivos se le llama **modelo lin-log**.

Interpretemos el coeficiente de la pendiente $\beta_2$.[17] Como de costumbre,

$$\beta_2 = \frac{\text{cambio en } Y}{\text{cambio en } \ln X} = \frac{\text{cambio en } Y}{\text{cambio relativo en } X}$$

El segundo paso se deriva de que *un cambio en el log de un número es un cambio relativo*.

Simbólicamente, tenemos

$$\beta_2 = \frac{\Delta Y}{\Delta X/X} \tag{6.6.12}$$

donde, como es usual, $\Delta$ denota un cambio pequeño. La ecuación (6.6.12) se escribe, en forma equivalente, así:

$$\Delta Y = \beta_2(\Delta X/X) \tag{6.6.13}$$

Esta ecuación plantea que el cambio absoluto en $Y (= \Delta Y)$ es igual a la pendiente multiplicada por el cambio relativo en $X$. Si este último se multiplica por 100, entonces (6.6.13) da el cambio absoluto en $Y$ ocasionado por un cambio porcentual en $X$. Así, si $\Delta X/X$ cambia en 0.01 unidades (o 1%), el cambio absoluto en $Y$ es $0.01(\beta_2)$. Por tanto, si en una aplicación se encuentra que $\beta_2 = 500$, el cambio absoluto en $Y$ es $(0.01)(500) = 5.0$. Por consiguiente, cuando se utiliza MCO para estimar regresiones como en (6.6.11), se debe multiplicar el valor del coeficiente estimado de la pendiente por 0.01, o, lo que es lo mismo, dividirlo entre 100. *Si no tiene presente lo anterior, la interpretación en una aplicación será muy equivocada.*

La pregunta práctica es: ¿cuándo resulta útil un modelo lin-log como el (6.6.11)? Se ha encontrado una interesante aplicación en los así llamados **modelos de gasto Engel** [en honor del estadístico alemán Ernst Engel (1821-1896)] (véase el ejercicio 6.10). Engel postuló que "el gasto total que se dedica a los alimentos tiende a incrementarse en progresión aritmética, mientras que el gasto total aumenta en progresión geométrica".[18]

**EJEMPLO 6.5**

Como ejemplo del modelo lin-log, revise el ejemplo sobre gasto alimentario en India, ejemplo 3.2. Ahí ajustamos un modelo lineal en las variables, como primera aproximación. Pero si graficamos los datos, obtenemos el gráfico de la figura 6.5. Como esta figura indica, el gasto alimentario se incrementa más despacio, conforme el gasto total aumenta, lo cual quizá sustente la ley de Engel. Los resultados de ajustar el modelo lin-log a los datos son los siguientes:

$$\begin{aligned} \widehat{\text{GasAli}}_i &= -1\,283.912 \quad + 257.2700 \ln \text{GasTot}_i \\ t &= \quad (-4.3848)^* \qquad (5.6625)^* \qquad r^2 = 0.3769 \end{aligned} \tag{6.6.14}$$

*Nota:* (*) denota un valor $p$ en extremo pequeño.

(*continúa*)

[17] Una vez más, mediante cálculo diferencial, tenemos

$$\frac{dY}{dX} = \beta_2\left(\frac{1}{X}\right)$$

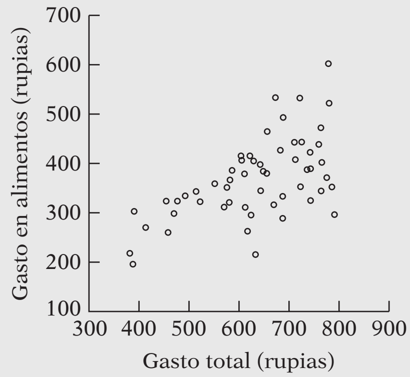
Por consiguiente,

$$\beta_2 = \frac{dY}{\frac{dX}{X}} = (6.6.12)$$

[18] Véase Chandan Mukherjee, Howard White y Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, Londres, 1998, p. 158. Esta cita se atribuye a H. Working, "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, vol. 38, 1943, pp. 43-56.

**EJEMPLO 6.5**
(*continuación*)

**FIGURA 6.5**




Interpretado de la forma antes descrita, el coeficiente de la pendiente, que vale casi 257, significa que un incremento en el gasto total en alimentos de 1%, en promedio, propicia un incremento de casi 2.57 rupias en el gasto en alimento de las 55 familias incluidas en la muestra. (*Nota:* El coeficiente estimado de la pendiente se dividió entre 100.)

Antes de seguir, observe que si desea calcular el coeficiente de elasticidad para los modelos lin-log o log-lin, puede hacerlo a partir de la definición del coeficiente de elasticidad dada antes, a saber:

$$\text{Elasticidad} = \frac{dY}{dX}\frac{X}{Y}$$

En la práctica, al conocer la forma funcional de un modelo, se pueden calcular las elasticidades con la definición anterior. (La tabla 6.6, que se presenta más adelante, resume los coeficientes de elasticidad para los diversos modelos.)

Debe señalarse que, algunas veces, la transformación logarítmica se emplea para reducir la heteroscedasticidad, así como la asimetría. (Véase el capítulo 11.) Una característica común de muchas variables económicas es que tienen asimetría positiva (por ejemplo, distribución del tamaño de las empresas, o distribución del ingreso o riqueza) y son heteroscedásticas. Una transformación logarítmica de dichas variables reduce tanto la asimetría como la heteroscedasticidad. Por eso, los economistas laborales acostumbran usar logaritmos de los salarios en la regresión de éstos sobre, por poner un ejemplo, el nivel de escolaridad, medido éste por los años de educación recibida.

## 6.7 Modelos recíprocos

Los modelos del siguiente tipo se conocen como modelos **recíprocos**.

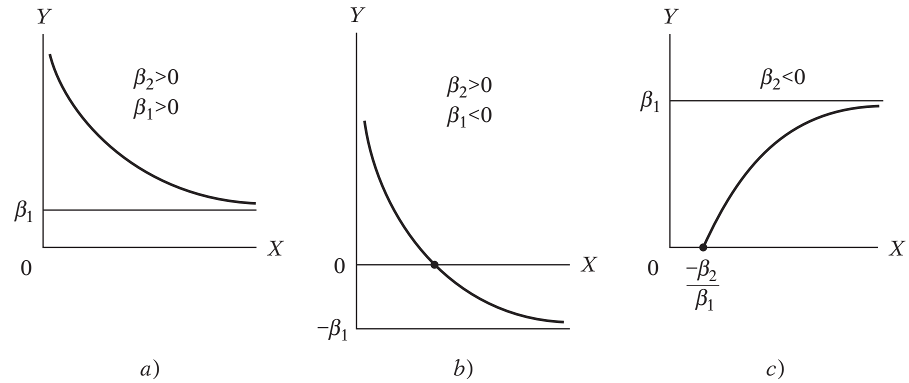
$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \tag{6.7.1}$$

A pesar de que este modelo es no lineal en la variable $X$ porque entra inversamente o en forma recíproca, el modelo es lineal en $\beta_1$ y $\beta_2$, y, por consiguiente, es un modelo de regresión lineal.[19]

Este modelo tiene las siguientes características: a medida que $X$ aumenta indefinidamente, el término $\beta_2\,(1/X)$ se acerca a cero (*nota:* $\beta_2$ es una constante) y $Y$ se aproxima al valor límite

[19] Si $X_i^* = (1/X_i)$, entonces la ecuación (6.7.1) es lineal en los parámetros al igual que en las variables $Y_i$ y $X_i^*$.

**FIGURA 6.6**
El modelo recíproco:
$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$.



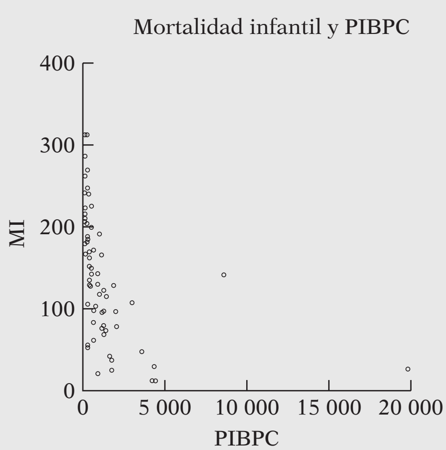

o *asintótico* $\beta_1$. Por consiguiente, modelos como (6.7.1) contienen un valor **asintótico** o límite que tomará la variable dependiente cuando el valor de la variable $X$ aumente indefinidamente.[20] Algunas formas probables de la curva correspondiente a (6.7.1) se muestran en la figura 6.6.

**EJEMPLO 6.6**

Como un ejemplo de la figura 6.6*a*, considere los datos de la tabla 6.4. Son datos transversales de 64 países sobre mortalidad infantil y otras variables. Por el momento, hay que concentrarse en las variables de mortalidad infantil (MI) y PIB *per cápita* (PIBPC), que se grafican en la figura 6.7.

Como se ve, esta figura se parece a la 6.6*a*: conforme se incrementa el PIB *per cápita*, se esperaría que la mortalidad infantil disminuyera, debido a que las personas pueden gastar más en el cuidado de la salud, si se supone que los demás factores permanecen constantes. Pero la relación no da como resultado una línea recta: conforme se incrementa el PIB *per cápita*, al principio existe una disminución drástica de la MI, pero disminuye en la medida en que el PIB *per cápita* sigue en aumento.

**FIGURA 6.7**
Relación entre la mortalidad infantil y el PIB *per cápita*, en 66 países.




(*continúa*)

[20] La pendiente de la ecuación (6.7.1) es: $dY/dX = -\beta_2(1/X^2)$, e implica que si $\beta_2$ es positivo, la pendiente siempre es negativa, y si $\beta_2$ es negativa, la pendiente siempre es positiva. Véanse las figuras 6.6*a* y 6.6*c*, respectivamente.

**EJEMPLO 6.6** (*continuación*)

**TABLA 6.4 Fecundidad y otros datos de 64 países**

| Observación | MI | TAF | PIBPC | TFT | Observación | MI | TAF | PIBPC | TFT |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 128 | 37 | 1 870 | 6.66 | 33 | 142 | 50 | 8 640 | 7.17 |
| 2 | 204 | 22 | 130 | 6.15 | 34 | 104 | 62 | 350 | 6.60 |
| 3 | 202 | 16 | 310 | 7.00 | 35 | 287 | 31 | 230 | 7.00 |
| 4 | 197 | 65 | 570 | 6.25 | 36 | 41 | 66 | 1 620 | 3.91 |
| 5 | 96 | 76 | 2 050 | 3.81 | 37 | 312 | 11 | 190 | 6.70 |
| 6 | 209 | 26 | 200 | 6.44 | 38 | 77 | 88 | 2 090 | 4.20 |
| 7 | 170 | 45 | 670 | 6.19 | 39 | 142 | 22 | 900 | 5.43 |
| 8 | 240 | 29 | 300 | 5.89 | 40 | 262 | 22 | 230 | 6.50 |
| 9 | 241 | 11 | 120 | 5.89 | 41 | 215 | 12 | 140 | 6.25 |
| 10 | 55 | 55 | 290 | 2.36 | 42 | 246 | 9 | 330 | 7.10 |
| 11 | 75 | 87 | 1 180 | 3.93 | 43 | 191 | 31 | 1 010 | 7.10 |
| 12 | 129 | 55 | 900 | 5.99 | 44 | 182 | 19 | 300 | 7.00 |
| 13 | 24 | 93 | 1 730 | 3.50 | 45 | 37 | 88 | 1 730 | 3.46 |
| 14 | 165 | 31 | 1 150 | 7.41 | 46 | 103 | 35 | 780 | 5.66 |
| 15 | 94 | 77 | 1 160 | 4.21 | 47 | 67 | 85 | 1 300 | 4.82 |
| 16 | 96 | 80 | 1 270 | 5.00 | 48 | 143 | 78 | 930 | 5.00 |
| 17 | 148 | 30 | 580 | 5.27 | 49 | 83 | 85 | 690 | 4.74 |
| 18 | 98 | 69 | 660 | 5.21 | 50 | 223 | 33 | 200 | 8.49 |
| 19 | 161 | 43 | 420 | 6.50 | 51 | 240 | 19 | 450 | 6.50 |
| 20 | 118 | 47 | 1 080 | 6.12 | 52 | 312 | 21 | 280 | 6.50 |
| 21 | 269 | 17 | 290 | 6.19 | 53 | 12 | 79 | 4 430 | 1.69 |
| 22 | 189 | 35 | 270 | 5.05 | 54 | 52 | 83 | 270 | 3.25 |
| 23 | 126 | 58 | 560 | 6.16 | 55 | 79 | 43 | 1 340 | 7.17 |
| 24 | 12 | 81 | 4 240 | 1.80 | 56 | 61 | 88 | 670 | 3.52 |
| 25 | 167 | 29 | 240 | 4.75 | 57 | 168 | 28 | 410 | 6.09 |
| 26 | 135 | 65 | 430 | 4.10 | 58 | 28 | 95 | 4 370 | 2.86 |
| 27 | 107 | 87 | 3 020 | 6.66 | 59 | 121 | 41 | 1 310 | 4.88 |
| 28 | 72 | 63 | 1 420 | 7.28 | 60 | 115 | 62 | 1 470 | 3.89 |
| 29 | 128 | 49 | 420 | 8.12 | 61 | 186 | 45 | 300 | 6.90 |
| 30 | 27 | 63 | 19 830 | 5.23 | 62 | 47 | 85 | 3 630 | 4.10 |
| 31 | 152 | 84 | 420 | 5.79 | 63 | 178 | 45 | 220 | 6.09 |
| 32 | 224 | 23 | 530 | 6.50 | 64 | 142 | 67 | 560 | 7.20 |

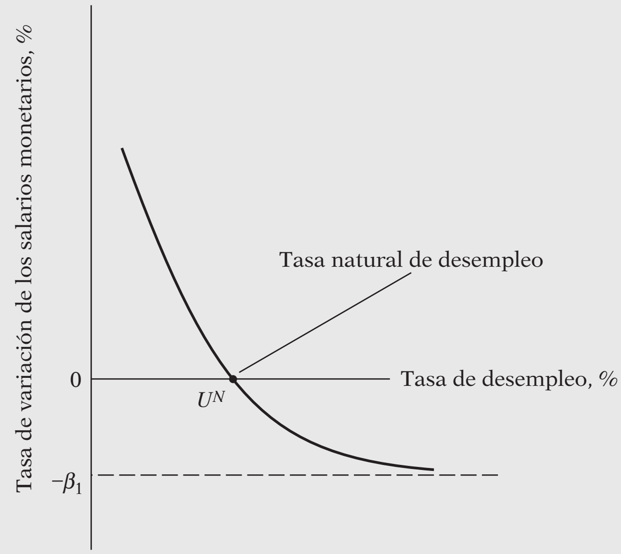
*Notas:* MI = mortalidad infantil, el número de defunciones de niños menores de 5 años en un año por cada 1 000 nacidos vivos.
TAF = tasa de alfabetismo femenina (porcentaje).
PIBPC = PIB *per cápita* en 1980.
TFT = tasa de fecundidad total, 1980-1985, cantidad promedio de hijos por mujer, con tasas de fecundidad para edades específicas en un año determinado.
Fuente: Chandan Mukherjee, Howard White y Marc Whyte, *Econometrics and Data Analysis for Developing Countries*, Routledge, Londres, 1998, p. 456.

Si tratamos de ajustar el modelo recíproco (6.7.1), obtenemos los siguientes resultados:

$$\begin{aligned} \widehat{\text{MI}}_i &= 81.79436 + 27\,237.17\left(\frac{1}{\text{PIBPC}_i}\right) \\ \text{ee} &= (10.8321) \qquad (3\,759.999) \\ t &= (7.5511) \qquad (7.2535) \qquad r^2 = 0.4590 \end{aligned} \tag{6.7.2}$$

Conforme el PIB *per cápita* se incrementa indefinidamente, la mortalidad infantil se acerca a su valor asintótico de casi 82 muertes por millar. Como se explicó en la nota 20, el valor positivo del coeficiente de ($1/\text{PIBPC}_t$) implica que la tasa de cambio de la MI respecto del PIBPC es negativa.

**FIGURA 6.8** La curva de Phillips.




Una aplicación importante de la figura 6.6*b* es la conocida curva de Phillips de macroeconomía. Con base en los datos de tasa de variación porcentual de los salarios nominales ($Y$) y la tasa porcentual de desempleo ($X$) en el Reino Unido durante el periodo 1861 a 1957, Phillips obtuvo una curva cuya forma general se parece a la figura 6.6*b* (figura 6.8).[21]

Como muestra la figura 6.8, existe asimetría en la respuesta de los cambios salariales en el nivel de la tasa de desempleo: los salarios aumentan con mayor rapidez por cada unidad de cambio en el desempleo si la tasa de desempleo está por debajo de $U^N$, denominada por los economistas *tasa natural de desempleo* (que se define como la tasa de desempleo requerida para mantener constante la inflación [salarial]), y luego disminuyen despacio por un cambio equivalente cuando la tasa de desempleo está por encima del nivel natural, $U^N$, lo que indica el piso asintótico, o $-\beta_1$, para el cambio salarial. Esta característica particular de la curva de Phillips puede deberse a factores institucionales, como el poder de negociación de los sindicatos, los salarios mínimos, compensaciones por desempleo, etcétera.

Desde la publicación del artículo de Phillips se efectuó una muy extensa investigación sobre la curva de Phillips tanto en el nivel teórico como en el empírico. El espacio de este libro no permite estudiar los detalles de la controversia en torno a la curva de Phillips. La curva misma ha pasado por diversas representaciones. Una formulación comparativamente reciente la proporciona Olivier Blanchard.[22] Si $\pi_t$ denota la tasa de inflación en el tiempo $t$, que se define como el cambio porcentual en el nivel de precios medido por un índice de precios representativo, como el índice de precios al consumidor (IPC), y si $UN_t$ denota la tasa de desempleo en el tiempo $t$, entonces la versión moderna de la curva de Phillips se expresa según el siguiente formato:

$$\pi_t - \pi_t^e = \beta_2(UN_t - U^N) + u_t \qquad \textbf{(6.7.3)}$$

donde $\pi_t =$ tasa real de inflación en el tiempo $t$

$\pi_t^e =$ tasa de inflación esperada en el tiempo $t$, donde la expectativa se forma en el año $(t-1)$

(*continúa*)

[21] A. W. Phillips, "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957", *Economica*, noviembre de 1958, vol. 25, pp. 283-299. Observe que la curva original no cruzó el eje de la tasa de desempleo, pero la figura 6.8 representa una versión posterior de la curva.

[22] Véase Olivier Blanchard, *Macroeconomics*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1997, capítulo 17.

**EJEMPLO 6.6**
(*continuación*)

$UN_t$ = tasa real de desempleo vigente en el tiempo $t$
$U^N$ = tasa natural de desempleo
$u_t$ = término de error estocástico[23]

Como $\pi_t^e$ no se puede observar de manera directa, en primer lugar se simplifica con la suposición de que $\pi_t^e = \pi_{t-1}$; es decir, la inflación esperada este año es la tasa de inflación que prevaleció el año anterior; por supuesto, se pueden formular suposiciones más complicadas respecto de la formación de expectativas, lo cual analizaremos en el capítulo 17, en los modelos de retraso distribuido.

Al sustituir esta suposición en (6.7.3) y escribir el modelo de regresión en la forma estándar, obtenemos la siguiente ecuación de estimación:

$$\pi_t - \pi_{t-1} = \beta_1 + \beta_2 UN_t + u_t \tag{6.7.4}$$

donde $\beta_1 = -\beta_2 U^N$. La ecuación (6.7.4) establece que el cambio en la tasa de inflación entre los dos periodos está linealmente relacionado con la tasa de desempleo real. *A priori*, se espera que $\beta_2$ sea negativa (¿por qué?) y $\beta_1$ positiva (porque $\beta_2$ es negativa y $U^N$ es positiva).

A propósito, la relación de Phillips de (6.7.3) se conoce en la bibliografía como la **curva de Phillips modificada**, **curva de Phillips de expectativas aumentadas** (para indicar que $\pi_{t-1}$ representa la inflación esperada) o **curva aceleradora de Phillips** (para indicar que una tasa de desempleo baja propicia un incremento en la tasa de inflación y, por consiguiente, una *aceleración* del nivel de precios).

**EJEMPLO 6.7**

A manera de ilustración de la curva de Phillips modificada, en la tabla 6.5 se presentan datos sobre la inflación medida por el porcentaje anual en el índice de precios al consumidor (IPCon inflación) y la tasa de desempleo de 1960 a 2006. La tasa de desempleo representa la tasa de desempleo civil. Con estos datos se obtuvo el cambio en la tasa de inflación ($\pi_t - \pi_{t-1}$) y se graficó respecto de la tasa de desempleo civil; se utiliza el IPC como medida de la inflación. La gráfica resultante aparece en la figura 6.9.

Como se esperaba, la relación entre el cambio en la tasa de inflación y la tasa de desempleo es negativa (un desempleo bajo provoca un incremento en la tasa de inflación y, por consiguiente, una aceleración del nivel de precios, de ahí el nombre de curva aceleradora de Phillips).

Al observar la figura 6.9, no resulta obvio si un modelo de regresión lineal (una línea recta) o un modelo recíproco sea el que se ajuste a los datos; tal vez haya una relación curvilínea entre las dos variables. Más adelante se presentan regresiones basadas en ambos modelos. Sin embargo, hay que tener presente que para el modelo recíproco se espera que el término del intercepto sea negativo y la pendiente positiva, como se destacó en la nota 20.

Modelo lineal:

$$\begin{aligned}\widehat{(\pi_t - \pi_{t-1})} &= 3.7844 \quad - \quad 0.6385\, UN_t \\ t &= (4.1912) \quad (-4.2756) \qquad r^2 = 0.2935\end{aligned} \tag{6.7.5}$$

Modelo recíproco:

$$\begin{aligned}\widehat{(\pi_t - \pi_{t-1})} &= -3.0684 \; + 17.2077\left(\frac{1}{UN_t}\right) \\ t &= (-3.1635) \qquad (3.2886) \qquad r^2 = 0.1973\end{aligned} \tag{6.7.6}$$

Todos los coeficientes estimados en ambos modelos son significativos estadística e *individualmente*; además, todos los valores $p$ son menores que el nivel 0.005.

[23] Los economistas creen que este término de error representa algún tipo de choque en la oferta, como los embargos de petróleo de la OPEP en 1973 y 1979.

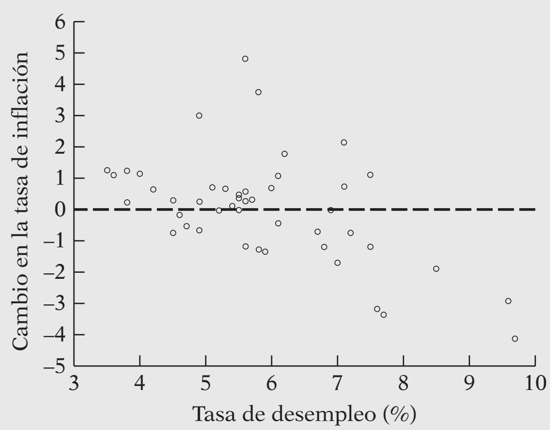
**TABLA 6.5**
**Tasas de inflación y desempleo, Estados Unidos, 1960-2006 (para todos los consumidores urbanos; 1982-1984 = 100, salvo si se indica lo contrario)**

Fuente: *Economic Report of the President*, 2007, tabla B-60, p. 399, para los cambios en el IPC, y tabla B-42, p. 376, para la tasa de desempleo.

| Año | TSINF | TSDES | Año | TSINF | TSDES |
|---|---|---|---|---|---|
| 1960 | 1.718 | 5.5 | 1984 | 4.317 | 7.5 |
| 1961 | 1.014 | 6.7 | 1985 | 3.561 | 7.2 |
| 1962 | 1.003 | 5.5 | 1986 | 1.859 | 7.0 |
| 1963 | 1.325 | 5.7 | 1987 | 3.650 | 6.2 |
| 1964 | 1.307 | 5.2 | 1988 | 4.137 | 5.5 |
| 1965 | 1.613 | 4.5 | 1989 | 4.818 | 5.3 |
| 1966 | 2.857 | 3.8 | 1990 | 5.403 | 5.6 |
| 1967 | 3.086 | 3.8 | 1991 | 4.208 | 6.8 |
| 1968 | 4.192 | 3.6 | 1992 | 3.010 | 7.5 |
| 1969 | 5.460 | 3.5 | 1993 | 2.994 | 6.9 |
| 1970 | 5.722 | 4.9 | 1994 | 2.561 | 6.1 |
| 1971 | 4.381 | 5.9 | 1995 | 2.834 | 5.6 |
| 1972 | 3.210 | 5.6 | 1996 | 2.953 | 5.4 |
| 1973 | 6.220 | 4.9 | 1997 | 2.294 | 4.9 |
| 1974 | 11.036 | 5.6 | 1998 | 1.558 | 4.5 |
| 1975 | 9.128 | 8.5 | 1999 | 2.209 | 4.2 |
| 1976 | 5.762 | 7.7 | 2000 | 3.361 | 4.0 |
| 1977 | 6.503 | 7.1 | 2001 | 2.846 | 4.7 |
| 1978 | 7.591 | 6.1 | 2002 | 1.581 | 5.8 |
| 1979 | 11.350 | 5.8 | 2003 | 2.279 | 6.0 |
| 1980 | 13.499 | 7.1 | 2004 | 2.663 | 5.5 |
| 1981 | 10.316 | 7.6 | 2005 | 3.388 | 5.1 |
| 1982 | 6.161 | 9.7 | 2006 | 3.226 | 4.6 |
| 1983 | 3.212 | 9.6 | | | |

*Nota:* La tasa de inflación es el cambio porcentual anual en el IPC. La tasa de desempleo es la tasa de desempleo civil.

**FIGURA 6.9**
Curva de Phillips modificada.

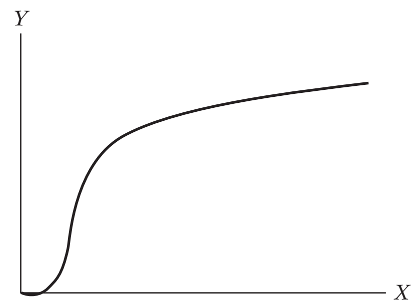



El modelo (6.7.5) muestra que si la tasa de desempleo baja un punto porcentual, en promedio, el cambio en la tasa de inflación aumenta 0.64 puntos porcentuales, y viceversa. El modelo (6.7.6) revela que, aunque la tasa de desempleo se incrementara de manera indefinida, el máximo cambio en la tasa de inflación bajaría y sería de 3.07 puntos porcentuales. A propósito, de la ecuación (6.7.5) se puede calcular la tasa de desempleo natural subyacente, de la siguiente forma:

$$U^N = \frac{\hat{\beta}_1}{-\hat{\beta}_2} = \frac{3.7844}{0.6385} = 5.9270 \tag{6.7.7}$$

Es decir, la tasa de desempleo natural es de casi 5.93%. Los economistas sitúan la tasa natural entre 5 y 6%, aunque en años recientes la tasa real de desempleo en Estados Unidos ha sido mucho más baja.

**FIGURA 6.10**
Modelo recíproco logarítmico.




## Modelo log hipérbola o recíproco logarítmico

Concluimos este análisis de los modelos recíprocos con el modelo recíproco logarítmico, que adopta la siguiente forma:

$$\ln Y_i = \beta_1 - \beta_2 \left(\frac{1}{X_i}\right) + u_i \tag{6.7.8}$$

Su forma se ilustra en la figura 6.10. Como se muestra ahí, al principio *Y* se incrementa con una tasa creciente (es decir, la curva es convexa al inicio) y luego aumenta con una tasa decreciente (la curva se convierte en cóncava).[24] Por consiguiente, este modelo sería apropiado para representar una función de producción de corto plazo. Recuerde que la microeconomía establece que si el trabajo y el capital son insumos en una función de producción, y si se mantiene constante el insumo capital pero se incrementa el insumo mano de obra, la relación entre producto y mano de obra de corto plazo se parecerá a la figura 6.10 (véase el ejemplo 7.4 del capítulo 7).

# 6.8 Elección de la forma funcional

En este capítulo se analizaron las distintas formas funcionales que un modelo empírico puede tomar, incluso dentro de los confines de los modelos de regresión lineal en los parámetros. La elección de una forma funcional particular puede ser relativamente fácil para el caso de dos variables, pues se pueden graficar las variables y tener así una ligera idea respecto del modelo adecuado. La elección se complica mucho más cuando se considera el modelo de regresión múltiple que implica más de una regresora, lo cual veremos al analizar este tema en los siguientes

[24] Con el cálculo, se demuestra que

$$\frac{d}{dX}(\ln Y) = -\beta_2\left(-\frac{1}{X^2}\right) = \beta_2\left(\frac{1}{X^2}\right)$$

Pero

$$\frac{d}{dX}(\ln Y) = \frac{1}{Y}\frac{dY}{dX}$$

Al sustituir, obtenemos

$$\frac{dY}{dX} = \beta_2\frac{Y}{X^2}$$

que es la pendiente de *Y* respecto de *X*.

**TABLA 6.6**

| Modelo | Ecuación | Pendiente $\left(=\frac{dY}{dX}\right)$ | Elasticidad $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$ |
|---|---|---|---|
| Lineal | $Y=\beta_1+\beta_2X$ | $\beta_2$ | $\beta_2\left(\frac{X}{Y}\right)^*$ |
| Log-lineal | $\ln Y=\beta_1+\beta_2\ln X$ | $\beta_2\left(\frac{Y}{X}\right)$ | $\beta_2$ |
| Log-lin | $\ln Y=\beta_1+\beta_2 X$ | $\beta_2\,(Y)$ | $\beta_2\,(X)^*$ |
| Lin-log | $Y=\beta_1+\beta_2\ln X$ | $\beta_2\left(\frac{1}{X}\right)$ | $\beta_2\left(\frac{1}{Y}\right)^*$ |
| Recíproco | $Y=\beta_1+\beta_2\left(\frac{1}{X}\right)$ | $-\beta_2\left(\frac{1}{X^2}\right)$ | $-\beta_2\left(\frac{1}{XY}\right)^*$ |
| Recíproco log | $\ln Y=\beta_1-\beta_2\left(\frac{1}{X}\right)$ | $\beta_2\left(\frac{Y}{X^2}\right)$ | $\beta_2\left(\frac{1}{X}\right)^*$ |

*Nota:* * indica que la elasticidad es variable: depende del valor tomado por $X$ o por $Y$, o por ambas. En la práctica, cuando no se especifican los valores de $X$ y de $Y$, es muy frecuente medir estas elasticidades con los valores medios de estas variables, es decir, $\bar{X}$ y $\bar{Y}$.

dos capítulos. No se niega que se requiere una gran destreza y experiencia al escoger un modelo apropiado para la estimación empírica. No obstante, se pueden mencionar algunas sugerencias.

1. La teoría (por ejemplo, la curva de Phillips) tal vez sugiera una forma funcional particular.

2. Es una buena costumbre calcular la tasa de cambio (es decir, la pendiente) de la regresada respecto de la regresora, así como conocer la elasticidad de la regresada respecto de la regresora. Para los diversos modelos estudiados en este capítulo, en la tabla 6.6 se ofrecen las fórmulas necesarias para los coeficientes de la pendiente y la elasticidad de los distintos modelos. Conocer estas fórmulas le servirá para comparar los diversos modelos.

3. Los coeficientes del modelo escogido deberán satisfacer determinadas expectativas *a priori*. Por ejemplo, si consideramos la demanda de automóviles como función del precio y otras variables, debemos esperar un coeficiente negativo para la variable precio.

4. Algunas veces, más de un modelo puede ajustarse razonablemente bien a un determinado conjunto de datos. En la curva de Phillips modificada, un modelo lineal y otro recíproco se ajustaron a los datos. En ambos casos, los coeficientes resultaron adecuados para las expectativas previas y fueron estadísticamente significativos. Una gran diferencia fue que el valor $r^2$ del modelo lineal fue mayor que el del modelo recíproco. Por tanto, se puede tener una ligera preferencia por el modelo lineal en comparación con el recíproco. *Pero se debe asegurar de que, al comparar dos valores de $r^2$, la variable dependiente (o regresada) de los dos modelos sea la misma; la(s) regresora(s) pueden tomar cualquier forma.* En el siguiente capítulo veremos por qué.

5. En general, *no se debe sobrevaluar* la medida de $r^2$ en el sentido de creer que mientras más alta sea $r^2$ mejor será el modelo. Como analizaremos en el siguiente capítulo, $r^2$ se incrementa conforme se añaden más regresoras al modelo. Lo que reviste mayor importancia es la justificación teórica del modelo elegido, los signos de los coeficientes estimados y su importancia estadística. Si un modelo es bueno conforme a estos criterios, quizá resulte aceptable un modelo con una $r^2$ menor. Examinaremos de nuevo este importante tema con mayor profundidad en el capítulo 13.

6. En algunas situaciones tal vez no sea fácil ponerse de acuerdo sobre una forma funcional concreta, en cuyo caso se pueden usar las llamadas transformaciones Box-Cox. En vista de que este tema es muy técnico, analizamos el procedimiento Box-Cox en el apéndice 6A.5.

## *6.9 Nota sobre la naturaleza del término de error estocástico: término de error estocástico aditivo o multiplicativo

Considere el siguiente modelo de regresión, similar a (6.5.1) pero sin el término de error:

$$Y_i = \beta_1 X_i^{\beta_2} \tag{6.9.1}$$

Para fines de estimación, este modelo se expresa de tres formas:

$$Y_i = \beta_1 X_i^{\beta_2} u_i \tag{6.9.2}$$

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.9.3}$$

$$Y_i = \beta_1 X_i^{\beta_2} + u_i \tag{6.9.4}$$

Al tomar logaritmos de ambos lados de estas ecuaciones, obtenemos

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i \tag{6.9.2a}$$

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.9.3a}$$

$$\ln Y_i = \ln \left(\beta_1 X_i^{\beta_2} + u_i\right) \tag{6.9.4a}$$

donde $\alpha = \ln \beta_1$.

Los modelos como (6.9.2) son de regresión *intrínsecamente lineales (en los parámetros)*, en el sentido de que, mediante una transformación (log) adecuada, los modelos se hacen lineales en los parámetros $\alpha$ y $\beta_2$. (*Nota:* Estos modelos son no lineales en $\beta_1$.) Pero el modelo (6.9.4) *intrínsecamente es no lineal en los parámetros*. No hay una manera simple de obtener el log de (6.9.4) porque $\ln (A + B) \neq \ln A + \ln B$.

A pesar de que las ecuaciones (6.9.2) y (6.9.3) son modelos de regresión lineal y se estiman por mínimos cuadrados ordinarios (MCO) o máxima verosimilitud (MV), se debe tener cuidado sobre las propiedades del término de error estocástico considerado en estos modelos. Recuerde que la propiedad MELI de MCO (mejor estimador lineal insesgado) exige que el valor de la media de $u_i$ sea cero, y que tenga varianza constante y autocorrelación cero. Para la prueba de hipótesis suponemos además que $u_i$ sigue una distribución normal con los valores de la media y la varianza recién estudiados. En resumen, supusimos que $u_i \sim N(0, \sigma^2)$.

Ahora considere el modelo (6.9.2). Su contraparte estadística está dada en (6.9.2*a*). Para utilizar el modelo clásico de regresión lineal normal (MCRLN) debemos suponer que

$$\ln u_i \sim N(0, \sigma^2) \tag{6.9.5}$$

Por consiguiente, cuando realicemos la regresión (6.9.2*a*), se tendrán que aplicar las pruebas de normalidad estudiadas en el capítulo 5 a los residuos obtenidos de esta regresión. A propósito, observe que si $\ln u_i$ sigue la distribución normal con media cero y varianza constante, la teoría estadística muestra que $u_i$ en (6.9.2) debe seguir la **distribución log-normal** con media $e^{\sigma^2/2}$ y varianza $e^{\sigma^2}(e^{\sigma^2} - 1)$.

Como muestra el análisis anterior, se tiene que prestar mucha atención al término de error al transformar un modelo para el análisis de regresión. En cuanto a (6.9.4), se trata de un modelo de regresión *no lineal en los parámetros,* y deberá resolverse mediante algún procedimiento computacional iterativo. La estimación del modelo (6.9.3) no debe presentar ningún problema.

* Opcional.

Para resumir, se debe prestar atención al término de perturbación cuando se transforme un modelo para el análisis de regresión. De lo contrario, una aplicación a ciegas de MCO al modelo transformado no producirá un modelo con las propiedades estadísticas deseables.

## Resumen y conclusiones

En este capítulo se presentaron diversos aspectos detallados del modelo clásico de regresión lineal (MCRL).

1. Algunas veces, un modelo de regresión puede no contener un término de intercepto explícito. Estos modelos se conocen como **regresión a través del origen**. A pesar de que el álgebra requerida en la estimación de tales modelos es simple, se deben utilizar con cautela. En tales modelos, la suma de los residuos $\sum \hat{u}_i$ es diferente de cero; además, el $r^2$ calculado convencionalmente puede no tener significado. A menos que exista una sólida razón teórica, es mejor introducir el intercepto explícitamente en el modelo.
2. Las unidades y la escala en que se expresan la regresada y la(s) regresora(s) son muy importantes, pues la interpretación de los coeficientes de regresión depende de ellas en gran medida. En la investigación empírica, el encargado no sólo debe citar la fuente de los datos, sino también describir explícitamente la forma como se miden las variables.
3. Es de igual importancia la forma funcional de la relación entre la regresada y la(s) regresora(s). Algunas formas funcionales importantes estudiadas en este capítulo son *a*) el modelo log-lineal o de elasticidad constante, *b*) los modelos de regresión semilogarítmicos y *c)* los modelos recíprocos.
4. En el modelo log-lineal, la regresada y la(s) regresora(s) se expresan en forma logarítmica. El coeficiente de regresión asociado al log de una regresora se interpreta como la elasticidad de la regresada respecto de la regresora.
5. En el modelo semilog, la regresada o la(s) regresora(s) están en la forma de log. En el modelo semilogarítmico, en el cual la regresada es logarítmica y la regresora $X$ es tiempo, el coeficiente de la pendiente estimado (multiplicado por 100) mide la tasa de crecimiento (instantánea) de la regresada. Tales modelos son comunes para medir la tasa de crecimiento de muchos fenómenos económicos. En el modelo semilogarítmico, si la regresora es logarítmica, su coeficiente mide la tasa de cambio absoluta en la regresada por un cambio porcentual dado en el valor de la regresora.
6. En los modelos recíprocos, la regresada o la regresora se expresa en forma recíproca o inversa para denotar relaciones no lineales entre variables económicas, como en la conocida curva de Phillips.
7. Al seleccionar las diversas formas funcionales, debe prestarse gran atención al término de perturbación estocástica $u_i$. Como vimos en el capítulo 5, el MCRL supone explícitamente que el valor de la media del término de perturbación es cero y su varianza es constante (homoscedástica), y que no está correlacionado con la(s) regresora(s). Con estos supuestos, los estimadores de MCO son MELI. Además, según el MCRLN, los estimadores de MCO están también normalmente distribuidos. Por consiguiente, se debe verificar si estos supuestos se mantienen en la forma funcional escogida para el análisis empírico. Después de realizar la regresión, el investigador debe aplicar pruebas de diagnóstico, como la de normalidad, estudiada en el capítulo 5. Nunca sobra insistir en este punto, pues las pruebas de hipótesis clásicas, como la $t$, $F$ y $\chi^2$, dependen del supuesto de que las perturbaciones están normalmente distribuidas. Esto es en especial importante si la muestra es pequeña.
8. Si bien el análisis hasta ahora se ha limitado a modelos de regresión con dos variables, los siguientes capítulos mostrarán que, en muchos casos, la extensión a modelos de regresión múltiple sólo implica más álgebra sin introducir necesariamente más conceptos fundamentales. Por esta razón, es muy importante que el lector tenga un concepto claro del modelo de regresión de dos variables.

# EJERCICIOS

## Preguntas

6.1. Considere el modelo de regresión

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

donde $y_i = (Y_i - \bar{Y})$ y $x_i = (X_i - \bar{X})$. En este caso, la línea de regresión debe pasar a través del origen. ¿Cierto o falso? Muestre sus cálculos.

6.2. Con base en datos mensuales de enero de 1978 a diciembre de 1987 se obtuvieron los siguientes resultados de regresión:

$$\begin{aligned}
\hat{Y}_t &= \ 0.00681 \qquad + 0.75815X_t \\
\text{ee} &= (0.02596) \qquad (0.27009) \\
t &= (0.26229) \qquad (2.80700) \\
\text{valor } p &= (0.7984) \qquad (0.0186) \qquad r^2 = 0.4406 \\
\hat{Y}_t &= \ 0.76214X_t \\
\text{ee} &= (0.265799) \\
t &= (2.95408) \\
\text{valor } p &= (0.0131) \qquad\qquad\qquad r^2 = 0.43684
\end{aligned}$$

donde $Y$ = tasa mensual de rendimiento de las acciones comunes de Texaco, %, y $X$ = tasa mensual de rendimiento del mercado, %.*

*a*) ¿Cuál es la diferencia entre los dos modelos de regresión?

*b*) Con los resultados anteriores, ¿conservaría el término del intercepto en el primer modelo? ¿Por qué?

*c*) ¿Cómo interpretaría los coeficientes de la pendiente en los dos modelos?

*d*) ¿Cuál es la teoría en los dos modelos?

*e*) ¿Puede comparar los términos $r^2$ de los dos modelos? ¿Por qué?

*f*) El estadístico de normalidad de Jarque-Bera para el primer modelo en este problema es 1.1167 y para el segundo modelo es 1.1170. ¿Qué conclusiones puede obtener de estos estadísticos?

*g*) El valor $t$ del coeficiente de la pendiente en el modelo con intercepto cero es aproximadamente 2.95, mientras que con el intercepto presente tiene un valor aproximado de 2.81. ¿Puede explicar este resultado?

6.3. Considere el siguiente modelo de regresión:

$$\frac{1}{Y_i} = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i$$

*Nota:* Ni $X$ ni $Y$ asumen el valor cero.

*a*) ¿Es un modelo de regresión lineal?

*b*) ¿Cómo estimaría este modelo?

*c*) ¿Cuál es el comportamiento de $Y$ a medida que $X$ tiende a infinito?

*d*) ¿Puede dar un ejemplo de un caso en el cual sea adecuado un modelo de esta clase?

* Los datos originales se obtuvieron del disquete de datos de Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading, Massachusetts, 1991.

6.4. Considere el modelo log-lineal:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Grafique $Y$ en el eje vertical y $X$ en el horizontal. Trace las curvas que exhiben la relación entre $Y$ y $X$ cuando $\beta_2 = 1$, cuando $\beta_2 > 1$ y cuando $\beta_2 < 1$.

6.5. Considere los siguientes modelos:

$$\text{Modelo I:} \qquad Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\text{Modelo II:} \qquad Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$$

donde $Y^*$ y $X^*$ son variables estandarizadas. Demuestre que $\hat{\alpha}_2 = \hat{\beta}_2(S_x/S_y)$ y, de allí, *establezca que, a pesar de que los coeficientes de pendiente son independientes de un cambio de origen, no lo son de un cambio de escala*.

6.6. Considere los siguientes modelos:

$$\ln Y_i^* = \alpha_1 + \alpha_2 \ln X_i^* + u_i^*$$
$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

donde $Y_i^* = w_1 Y_i$ y $X_i^* = w_2 X_i$, con las $w$ constantes.

*a*) Establezca las relaciones entre los dos conjuntos de coeficientes de regresión y sus errores estándar.

*b*) ¿Es diferente el $r^2$ en los dos modelos?

6.7. Entre las regresiones (6.6.8) y (6.6.10), ¿qué modelo es preferible? ¿Por qué?

6.8. Para la regresión (6.6.8), pruebe la hipótesis de que el coeficiente de la pendiente no es significativamente diferente de 0.005.

6.9. De la curva de Phillips estimada dada en (6.7.3), ¿es posible estimar la tasa natural de desempleo? ¿Cómo?

6.10. La curva de gasto de Engel relaciona el gasto del consumidor sobre un bien con su ingreso total. Sea $Y =$ el gasto de consumo sobre un bien y $X =$ ingreso del consumidor, y considere los siguientes modelos:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 + \beta_2 (1/X_i) + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 (1/X_i) + u_i$$
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

¿Cuál(es) de estos modelo(s) escogería para la curva de gasto de Engel y por qué? (*Sugerencia:* Interprete los diversos coeficientes de pendiente, encuentre las expresiones para la elasticidad del gasto respecto del ingreso, etcétera.)

6.11. Considere el siguiente modelo:

$$Y_i = \frac{e^{\beta_1 + \beta_2 X_i}}{1 + e^{\beta_1 + \beta_2 X_i}}$$

Tal como se presenta, ¿es un modelo de regresión lineal? Si no es así, ¿qué "truco" podría utilizar, si acaso, para convertirlo en un modelo de regresión lineal? ¿Cómo interpretaría el modelo resultante? ¿En qué circunstancias sería adecuado dicho modelo?

6.12. Grafique los siguientes modelos (para mayor sencillez, se omitieron los subíndices de observación, $i$):

*a*) $Y = \beta_1 X^{\beta_2}$, para $\beta_2 > 1$, $\beta_2 = 1$, $0 < \beta_2 < 1, \ldots$

*b*) $Y = \beta_1 e^{\beta_2 X}$, para $\beta_2 > 0$ y $\beta_2 < 0$.

Analice dónde serían adecuados estos modelos.

6.13 Considere la siguiente regresión:*

$$\begin{aligned} \text{ISP}_i &= -17.8 + 33.2\ \text{Gini}_i \\ \text{ee} &= \quad (4.9)\quad (11.8) \qquad r^2 = 0.16 \end{aligned}$$

donde ISP = índice de inestabilidad sociopolítica, promedio para 1960-1985, y Gini = coeficiente de Gini para 1975, o el año más próximo disponible dentro del periodo 1970-1980. La muestra consta de 40 países.

El coeficiente de Gini es una medida de la desigualdad del ingreso y se sitúa entre 0 y 1. Cuanto más cerca se encuentre de 0, mayor será la igualdad del ingreso, y cuanto más próximo se encuentra de 1, mayor será la desigualdad del ingreso.

*a*) ¿Cómo interpreta esta regresión?

*b*) Suponga que el coeficiente de Gini se incrementa de 0.25 a 0.55. ¿Cuánto aumentaría el ISP? ¿Qué significa eso en la práctica?

*c*) ¿El coeficiente estimado de la pendiente es estadísticamente significativo al nivel de 5%? Realice los cálculos necesarios.

*d*) Con base en la regresión anterior, ¿diría que los países con mayor desigualdad del ingreso son políticamente inestables?

## Ejercicios empíricos

6.14. Con los datos de la tabla 6.7,** ajuste el siguiente modelo a dichos datos, obtenga las estadísticas usuales de regresión e interprete los resultados:

$$\frac{100}{100 - Y_i} = \beta_1 + \beta_2\left(\frac{1}{X_i}\right)$$

**TABLA 6.7**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $Y_i$ | 86 | 79 | 76 | 69 | 65 | 62 | 52 | 51 | 51 | 48 |
| $X_i$ | 3 | 7 | 12 | 17 | 25 | 35 | 45 | 55 | 70 | 120 |

6.15. Para estudiar la relación entre tasa de inversión (el gasto en inversión como razón del PNB) y la tasa de ahorro (el ahorro como razón del PNB), Martin Feldstein y Charles Horioka recopilaron datos para una muestra de 21 países. (Véase la tabla 6.8.) La tasa de inversión de cada país es la tasa promedio correspondiente al periodo 1960-1974, y la tasa de ahorro es la tasa de ahorro promedio para el periodo 1960-1974. La variable TASINV representa la tasa de inversión, y la variable TASAHO, la tasa de ahorro.†

*a*) Grafique la tasa de inversión contra la tasa de ahorro.

*b*) Con base en esta gráfica, ¿considera que los siguientes modelos puedan ajustarse a los datos igualmente bien?

$$\begin{aligned} \text{Tasinv}_i &= \beta_1 + \beta_2 \text{Tasaho}_i + u_i \\ \ln \text{Tasinv}_i &= \alpha_1 + \alpha_2 \ln \text{Tasaho}_i + u_i \end{aligned}$$

*c*) Estime estos dos modelos y obtenga las estadísticas habituales.

* Véase David N. Weil, *Economic Growth*, Addison-Wesley, Boston, 2005, p. 392.

** Adaptado de J. Johnston, *Econometric Methods*, 3a. ed., McGraw-Hill, Nueva York, 1984, p. 87. En realidad, esta información se tomó de un examen de econometría de la Universidad de Oxford, 1975.

† Martin Feldstein y Charles Horioka, "Domestic Saving and International Capital Flows", *Economic Journal*, vol. 90, junio de 1980, pp. 314-329. Datos reproducidos de Michael P. Murray, *Econometrics: A Modern Introduction*, Addison-Wesley, Boston, 2006.

**TABLA 6.8**

| | TASAHO | TASINV |
|---|---|---|
| Alemania | 0.271 | 0.264 |
| Australia | 0.250 | 0.270 |
| Austria | 0.285 | 0.282 |
| Bélgica | 0.235 | 0.224 |
| Canadá | 0.219 | 0.231 |
| Dinamarca | 0.202 | 0.224 |
| España | 0.235 | 0.241 |
| Estados Unidos | 0.186 | 0.186 |
| Finlandia | 0.288 | 0.305 |
| Francia | 0.254 | 0.260 |
| Grecia | 0.219 | 0.248 |
| Irlanda | 0.190 | 0.218 |
| Italia | 0.235 | 0.224 |
| Japón | 0.372 | 0.368 |
| Luxemburgo | 0.313 | 0.277 |
| Noruega | 0.278 | 0.299 |
| Nueva Zelanda | 0.232 | 0.249 |
| Países Bajos | 0.273 | 0.266 |
| Reino Unido | 0.184 | 0.192 |
| Suecia | 0.241 | 0.242 |
| Suiza | 0.297 | 0.297 |

*Nota:* TASAHO = Ahorro como razón del PIB.
TASINV = Gasto en inversión como razón del PIB.

*d*) ¿Cómo interpretaría el coeficiente de la pendiente en el modelo lineal? ¿Y en el modelo log-lineal? ¿Hay alguna diferencia en la interpretación de estos coeficientes?

*e*) ¿Cómo interpretaría los interceptos de los dos modelos? ¿Hay alguna diferencia en la interpretación?

*f*) ¿Compararía los dos coeficientes $r^2$? ¿Por qué?

*g*) Suponga que desea calcular la elasticidad de la tasa de inversión respecto de la tasa de ahorro. ¿Cómo obtendría esta elasticidad para el modelo lineal? ¿Y para el modelo log-lineal? Tenga en cuenta que esta elasticidad se define como el cambio porcentual de la tasa de inversión correspondiente a un cambio porcentual en la tasa de ahorro.

*h*) Con los resultados de los dos modelos de regresión, ¿qué modelo preferiría? ¿Por qué?

6.16. La tabla 6.9* presenta las definiciones de las variables para diversos tipos de gastos, gasto total, ingreso, edad del jefe de la familia y número de hijos para una muestra de 1 519 familias, que se obtuvo de los Estudios Económicos de Gasto de las Familias Británicas 1980-1982.

El conjunto original de datos se encuentra en la página de internet de este texto. Los datos incluyen sólo a familias con uno o dos hijos que viven en la zona metropolitana de Londres. La muestra no incluye a familias de personas que trabajan por su cuenta o jubiladas.

*a*) Con los datos sobre el gasto alimentario en relación con el gasto total, determine qué modelos de los que se resumieron en la tabla 6.6 se ajustan a estos datos.

*b*) Con base en los resultados de regresión obtenidos en *a*), ¿qué modelo parece el más apropiado en el presente caso?

*Nota:* Guarde estos datos para análisis posterior en el siguiente capítulo sobre regresión múltiple.

* Los datos son de Richard Blundell y Krishna Pendakur, "Semiparametric Estimation and Customer demand", *Journal of Applied Econometrics*, vol. 13, núm. 5, 1998, pp. 435-462. Los datos se reproducen de R. Carter Hill, William E. Griffiths y George G. Judge, *Undergraduate Econometrics*, 2a. ed., John Wiley & Sons, Nueva York, 2001.

**TABLA 6.9**

Lista de variables:

*wfood* = participación del presupuesto destinado al gasto en alimentos
*wfuel* = participación del presupuesto destinado al gasto en combustibles
*wcloth* = participación del presupuesto destinado al gasto en ropa
*walc* = participación del presupuesto destinado al gasto en bebidas alcohólicas
*wtrans* = participación del presupuesto destinado al gasto en transporte
*wother* = participación del presupuesto destinado a otros gastos

*totexp* = gasto total familiar (redondeado a las 10 libras esterlinas más próximas)
*income* = ingreso total neto familiar (redondeado a las 10 libras esterlinas más próximas)
*age* = edad del jefe de la familia
*nk* = número de hijos

La participación del presupuesto destinado a un concepto, por ejemplo, alimentos, se define como:

$$wfood = \frac{\text{gasto en alimentos}}{\text{gasto total}}$$

6.17. Remítase a la tabla 6.3. Determine la tasa de crecimiento del gasto en bienes duraderos. ¿Cuál es la *semielasticidad* estimada? Interprete sus resultados. ¿Tendría sentido hacer una regresión doble logarítmica con el gasto en bienes duraderos como la variable dependiente (regresada) y el tiempo como la regresora? ¿Cómo interpretaría la pendiente del coeficiente en este caso?

6.18. A partir de los datos de la tabla 6.3, calcule la tasa de crecimiento del gasto en bienes perecederos y compare los resultados con los obtenidos en el ejercicio 6.17.

6.19. La tabla 6.10 presenta datos sobre el gasto de consumo total en el Reino Unido (en millones de £) y el gasto en publicidad (en millones de £) correspondientes a 29 categorías de producto.*

*a*) Considerando las diferentes formas funcionales que analizamos en el capítulo, ¿qué forma funcional puede ajustarse a los datos de la tabla 6.10?

*b*) Estime los parámetros del modelo de regresión seleccionado e interprete sus resultados.

*c*) Si toma la razón entre el gasto en publicidad y el gasto de consumo total, ¿qué observa? ¿Hay alguna categoría de producto en la que esta razón parezca extraordinariamente alta? ¿Estas categorías de producto tienen algo especial que explique el gasto relativamente alto en publicidad?

6.20. Remítase al ejemplo 3.3 del capítulo 3 para realizar lo siguiente:

*a*) Grafique la demanda de teléfonos celulares contra el ingreso *per cápita* ajustado por el poder adquisitivo (PA).

*b*) Grafique la demanda de teléfonos celulares contra el log del ingreso *per cápita* ajustado por el poder adquisitivo (PA).

*c*) ¿Qué diferencia hay entre las dos gráficas?

*d*) Con base en estas dos gráficas, ¿cree que un modelo doble logaritmo puede ofrecer un mejor ajuste a los datos que el modelo lineal? Estime el modelo doble-log.

*e*) ¿Cómo interpreta el coeficiente de la pendiente en el modelo doble-log?

*f*) ¿El coeficiente estimado de la pendiente en el modelo doble-log es estadísticamente significativo en el nivel de 5%?

* Estos datos se tomaron del *Advertising Statistics Year Book*, 1996, y se reproducen de http://www.Economicswebinstitute.org/ecdata.htm.

**TABLA 6.10**
**Gasto en publicidad y gasto de consumo total (en millones de £) en 29 categorías de productos en el Reino Unido**

Fuente: http://www.Economicswebinstitute.org/ecdata.htm.

| Obs | GASPUB | GASCON | RAZÓN |
|---|---|---|---|
| 1 | 87 957.00 | 13 599.00 | 0.006468 |
| 2 | 23 578.00 | 4 699.000 | 0.005018 |
| 3 | 16 345.00 | 5 473.000 | 0.002986 |
| 4 | 6 550.000 | 6 119.000 | 0.001070 |
| 5 | 10 230.00 | 8 811.000 | 0.001161 |
| 6 | 9 127.000 | 1 142.000 | 0.007992 |
| 7 | 1 675.000 | 143.0000 | 0.011713 |
| 8 | 1 110.000 | 138.0000 | 0.008043 |
| 9 | 3 351.000 | 85.00000 | 0.039424 |
| 10 | 1 140.000 | 108.0000 | 0.010556 |
| 11 | 6 376.000 | 307.0000 | 0.020769 |
| 12 | 4 500.000 | 1 545.000 | 0.002913 |
| 13 | 1 899.000 | 943.0000 | 0.002014 |
| 14 | 10 101.00 | 369.0000 | 0.027374 |
| 15 | 3 831.000 | 285.0000 | 0.013442 |
| 16 | 99 528.00 | 1 052.000 | 0.094608 |
| 17 | 15 855.00 | 862.0000 | 0.018393 |
| 18 | 8 827.000 | 84.00000 | 0.105083 |
| 19 | 54 517.00 | 1 174.000 | 0.046437 |
| 20 | 49 593.00 | 2 531.000 | 0.019594 |
| 21 | 39 664.00 | 408.0000 | 0.097216 |
| 22 | 327.0000 | 295.0000 | 0.001108 |
| 23 | 22 549.00 | 488.0000 | 0.046207 |
| 24 | 416 422.0 | 19 200.00 | 0.021689 |
| 25 | 14 212.00 | 94.00000 | 0.151191 |
| 26 | 54 174.00 | 5 320.000 | 0.010183 |
| 27 | 20 218.00 | 357.0000 | 0.056633 |
| 28 | 11 041.00 | 159.0000 | 0.069440 |
| 29 | 22 542.00 | 244.0000 | 0.092385 |

*Notas:* GASPUB = gasto en publicidad (en millones de £).
GASCON = Gasto de consumo total (en millones de £).

*g*) Cómo estimaría la elasticidad de la demanda de teléfonos celulares respecto del ingreso ajustado por el poder adquisitivo para el modelo lineal de la ecuación (3.7.3)? ¿Qué información adicional, si acaso, se necesita? Llame *elasticidad ingreso* a la elasticidad estimada.

*h*) ¿Hay alguna diferencia entre la elasticidad ingreso estimada con base en el modelo doble-log y la que se estimó con el modelo lineal? De ser así, ¿qué modelo elegiría?

6.21. Repita el ejercicio 6.20, pero remítase a la demanda de computadoras personales presentada en la ecuación (3.7.4). ¿Hay alguna diferencia entre la elasticidad ingreso estimada de los teléfonos celulares y la de las computadoras personales? De ser así, ¿qué factores pueden explicar la diferencia?

6.22. Consulte los datos de la tabla 3.3. Para averiguar si las personas con computadoras personales tienen también teléfonos celulares, ejecute la siguiente regresión:

$$\text{Telcelular}_i = \beta_1 + \beta_2 \text{PCs}_i + u_i$$

*a*) Estime los parámetros de esta regresión.

*b*) ¿El coeficiente estimado de la pendiente es estadísticamente significativo?

*c*) ¿Importa si se ejecuta la siguiente regresión?

$$\text{PCs}_i = \alpha_1 + \alpha_2 \text{Telcelular}_i + u_i$$

*d*) Estime la regresión anterior y pruebe la significancia estadística del coeficiente estimado de la pendiente.

*e*) ¿Cómo decidiría entre la primera y la segunda regresión?

# Apéndice 6A

## 6A.1 Derivación de los estimadores de mínimos cuadrados para la regresión a través del origen

Deseamos reducir

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2 \tag{1}$$

respecto de $\hat{\beta}_2$.

Al diferenciar (1) respecto de $\hat{\beta}_2$, obtenemos

$$\frac{d \sum \hat{u}_i^2}{d\hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_2 X_i)(-X_i) \tag{2}$$

Al igualar (2) a cero y simplificar, obtenemos

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (6.1.6) = (3)$$

Ahora sustituimos la FRP: $Y_i = \beta_2 X_i + u_i$ en esta ecuación, y obtenemos

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum X_i(\beta_2 X_i + u_i)}{\sum X_i^2} \\ &= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} \end{aligned} \tag{4}$$

[*Nota:* $E(\hat{\beta}_2) = \beta_2$.] Por consiguiente,

$$E(\hat{\beta}_2 - \beta_2)^2 = E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \tag{5}$$

Al expandir el lado derecho de (5) y observar que las $X_i$ son no estocásticas y las $u_i$ son homoscedásticas y no correlacionadas, tenemos

$$\operatorname{var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad (6.1.7) = (6)$$

A propósito, observe que a partir de (2) obtenemos, después de igualarla a cero,

$$\sum \hat{u}_i X_i = 0 \tag{7}$$

Del apéndice 3A, sección 3A.1, vemos que, cuando el término del intercepto está presente en el modelo, obtenemos, además de (7), la condición $\sum \hat{u}_i = 0$. Del desarrollo matemático anterior debe quedar clara la razón por la cual la regresión a través del origen puede no tener la suma de errores, $\sum \hat{u}_i$, igual a cero.

Suponga que deseamos imponer la condición $\sum \hat{u}_i = 0$. En ese caso, tenemos

$$\begin{aligned} \sum Y_i &= \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= \hat{\beta}_2 \sum X_i, \qquad \text{pues } \sum \hat{u}_i = 0 \text{ por construcción} \end{aligned} \tag{8}$$

Esta expresión entonces da

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum Y_i}{\sum X_i} \\ &= \frac{\bar{Y}}{\bar{X}} = \frac{\text{valor de la media de } Y}{\text{valor de la media de } X} \end{aligned} \tag{9}$$

Pero este estimador no es el mismo que el definido antes en (3) o en (6.1.6). Y como el $\hat{\beta}_2$ de (3) es insesgado (¿por qué?), el $\hat{\beta}_2$ de (9) no puede serlo.

El punto clave es que, en la regresión a través del origen, no es posible igualar $\sum \hat{u}_i X_i$ y $\sum \hat{u}_i$ a cero, como en el modelo convencional. La única condición que se satisface es que $\sum \hat{u}_i X_i$ es cero.

Recuerde que

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{2.6.3}$$

Al sumar esta ecuación en ambos lados y dividir entre $N$ el tamaño de la muestra, obtenemos

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \tag{10}$$

Como para el modelo de intercepto cero $\sum \hat{u}_i$ y, por consiguiente $\bar{\hat{u}}$, no necesariamente es cero, se cumple que

$$\bar{Y} \neq \bar{\hat{Y}} \tag{11}$$

es decir, la media de los valores observados de $Y$ no necesita ser igual a la media de los valores estimados de $Y$; los dos valores de las medias son idénticos en el modelo con intercepto, como se ve en (3.1.10).

Se observó que, para el modelo de intercepto cero, el $r^2$ puede ser negativo, mientras que para el modelo convencional, nunca puede ser negativo. Esta condición se demuestra de la siguiente manera:

Con (3.5.5*a*), escribimos

$$r^2 = 1 - \frac{\text{SCR}}{\text{SCT}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \tag{12}$$

Ahora, para el modelo convencional, o con presencia de intercepto, la ecuación (3.3.6) muestra que

$$\text{SCR} = \sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \leq \sum y_i^2 \tag{13}$$

a menos que $\hat{\beta}_2$ sea cero (es decir, $X$ no ejerce influencia alguna sobre $Y$). Dicho de otro modo, para el modelo convencional, SCR $\leq$ SCT, y así, el $r^2$ nunca podrá ser negativo.

De la misma forma, para el modelo de intercepto cero, se demuestra que

$$\text{SCR} = \sum \hat{u}_i^2 = \sum Y_i^2 - \hat{\beta}_2^2 \sum X_i^2 \tag{14}$$

(*Nota:* Las sumas de los cuadrados de $Y$ y $X$ no están ajustadas por la media.) Ahora, no existe garantía de que esta SCR sea siempre menor que $\sum y_i^2 = \sum Y_i^2 - N\bar{Y}^2$ (la SCT), lo cual indica que la SCR puede ser mayor que la SCT, y esto implica que el $r^2$, como se definió convencionalmente, puede ser negativo. A propósito, observe que, en este caso, la SCR será mayor que la SCT si $\hat{\beta}_2^2 \sum X_i^2 < N\bar{Y}^2$.

## 6A.2 Prueba de que la variable estandarizada tiene media cero y varianza unitaria

Considere la variable aleatoria (v.a.) $Y$ con valor medio (muestral) $\bar{Y}$ y desviación estándar (muestral) de $S_y$. Defina

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_y} \tag{15}$$

Por tanto, $Y_i^*$ es una variable estandarizada. Observe que la estandarización implica una operación dual: 1) el cambio del origen, que es el numerador de (15), y 2) la variación de escala, que es el denominador. En consecuencia, la estandarización implica una modificación en el origen y en la escala.

Ahora,

$$\bar{Y}_i^* = \frac{1}{S_y} \frac{\sum (Y_i - \bar{Y})}{n} = 0 \tag{16}$$

pues la suma de la desviación de una variable respecto del valor de su media siempre es cero. Así, el valor de la media del valor estandarizado es cero. (*Nota:* Podemos sacar el término $S_y$ del signo de la sumatoria porque su valor es conocido.)

Ahora,

$$\begin{aligned} S_{y^*}^2 &= \sum \frac{(Y_i - \bar{Y})^2/(n-1)}{S_y^2} \\ &= \frac{1}{(n-1)S_y^2} \sum (Y_i - \bar{Y})^2 \\ &= \frac{(n-1)S_y^2}{(n-1)S_y^2} = 1 \end{aligned} \tag{17}$$

Observe que

$$S_y^2 = \frac{\sum(Y_i - \bar{Y})^2}{n - 1}$$

lo cual es la varianza muestral de $Y$.

## 6A.3 Logaritmos

Considere los números 5 y 25. Sabemos que

$$25 = 5^2 \tag{18}$$

Decimos que el *exponente* 2 es el *logaritmo* de 25 con *base* 5. En términos más formales, el logaritmo de un número (por ejemplo, 25) con una base determinada (por ejemplo, 5) es la potencia (2) a la que debe elevarse la base (5) para obtener el número dado (25).

De manera más general, si

$$Y = b^X \, (b > 0) \tag{19}$$

entonces

$$\log_b Y = X \tag{20}$$

En matemáticas, la función (19) se llama *función exponencial,* y la función (20), *función logarítmica.* Como queda claro por las ecuaciones (19) y (20), una función es el inverso de la otra.

Aunque se puede usar cualquier base (positiva), en la práctica las dos bases más comunes son 10 y el número matemático $e = 2.71828\ldots$

Los logaritmos base 10 se llaman *logaritmos comunes.* Así,

$$\log_{10} 100 = 2 \quad \log_{10} 30 \approx 1.48$$

Es decir, en el primer caso, $100 = 10^2$, y en el segundo, $30 \approx 10^{1.48}$.

Los logaritmos base $e$ se llaman *logaritmos naturales.* Por tanto,

$$\log_e 100 \approx 4.6051 \text{ y } \log_e 30 \approx 3.4012$$

Todos estos cálculos suelen realizarse con una calculadora de mano.

Por convención, el logaritmo base 10 se denota por medio de las letras log, y el logaritmo base $e$, por ln. Así, en el ejemplo anterior, se puede escribir log 100 o log 30, o ln 100 o ln 30.

Existe una relación fija entre el log común y el log natural, que es

$$\ln X = 2.3026 \log X \tag{21}$$

Es decir, el logaritmo natural del número $X$ es igual a 2.3026 veces el logaritmo de $X$ base 10. Por consiguiente,

$$\ln 30 = 2.3026 \log 30 = 2.3026\,(1.48) = 3.4012 \text{ (aprox.)}$$

como antes. Por tanto, no importa si se usan logaritmos comunes o naturales. Sin embargo, en matemáticas, la base que casi siempre se prefiere es $e$, es decir, el logaritmo natural. En consecuencia, en este libro todos los logaritmos son naturales, a menos que expresamente se indique lo contrario. Por supuesto, se puede convertir el logaritmo de un número de una base a la otra con la ecuación (21).

Tenga presente que los logaritmos de los números negativos no están definidos. Por tanto, el logaritmo de $(-5)$ o el $\ln(-5)$ no está definido.

Algunas propiedades de los logaritmos son las siguientes: si $A$ y $B$ son números positivos, se demuestra que:

1. $$\ln(A \times B) = \ln A + \ln B \tag{22}$$

Es decir, el logaritmo del producto de dos números (positivos) $A$ y $B$ es igual a la suma de sus logaritmos.

2. $$\ln(A/B) = \ln A - \ln B \tag{23}$$

Es decir, el logaritmo de la razón de $A$ a $B$ es la diferencia entre los logaritmos de $A$ y $B$.

3. $$\ln(A \pm B) \neq \ln A \pm \ln B \tag{24}$$

Es decir, el logaritmo de la suma o diferencia de $A$ y $B$ no es igual a la suma o diferencia de sus logaritmos.

4. $$\ln (A^k) = k \ln A \tag{25}$$

Es decir, el logaritmo de $A$ elevado a la potencia $k$ es $k$ veces el logaritmo de $A$.

5. $$\ln e = 1 \tag{26}$$

Esto es, el logaritmo de $e$ base $e$ es 1 (lo mismo que el log de 10 base 10).

6. $$\ln 1 = 0 \tag{27}$$

Es decir, el logaritmo natural del número 1 es cero (al igual que el logaritmo común del número 1).

7. Si $Y = \ln X$,

$$\frac{dY}{dX} = \frac{1}{X} \tag{28}$$

Esto es, la tasa de cambio (es decir, la derivada) de $Y$ respecto de $X$ es 1 sobre $X$. Las funciones exponencial y logarítmica (natural) se muestran en la figura 6A.1.

Aunque el número cuyo logaritmo se toma es siempre positivo, el logaritmo de dicho número puede ser positivo o negativo. Se comprueba fácilmente que si

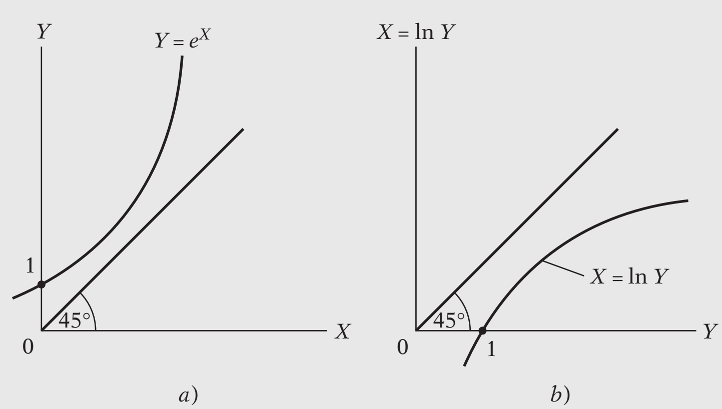
$$0 < Y < 1 \quad \text{entonces} \quad \ln Y < 0$$
$$Y = 1 \quad \text{entonces} \quad \ln Y = 0$$
$$Y > 1 \quad \text{entonces} \quad \ln Y > 0$$

Además, observe que, aunque la curva logarítmica que se ilustra en la figura 6A.1*b*) se inclina positivamente, lo que implica que cuanto más grande sea el número mayor será también el valor logarítmico, la curva se incrementa con una tasa decreciente (en matemáticas, la segunda derivada de la función es negativa). Así, $\ln(10) = 2.3026$ (aproximadamente) y $\ln(20) = 2.9957$ (aproximadamente). Esto es, si un número se duplica, su logaritmo no aumenta al doble.

Por esta razón, la transformación logarítmica se llama transformación no lineal. Esto también se aprecia en la ecuación (28), que indica que si $Y = \ln X$, $dY/dX = 1/X$. Esto significa que la pendiente de la función logarítmica depende del valor de $X$; es decir, no es constante (recuerde la definición de linealidad en la variable).

*Logaritmos y porcentajes:* Como $\frac{d(\ln X)}{dX} = \frac{1}{X}$ o $d(\ln X) = \frac{dX}{X}$, para cambios muy pequeños, el cambio en $\ln X$ es igual al cambio relativo o proporcional en $X$. En la práctica, si el cambio en $X$ es razonablemente pequeño, la relación anterior se escribe como el cambio en $\ln X \approx$ al cambio relativo en $X$, donde $\approx$ significa aproximadamente igual.

**FIGURA 6A.1**
Funciones exponencial y logarítmica: *a*) función exponencial; *b*) función logarítmica.



Así, para cambios pequeños,

$$(\ln X_t - \ln X_{t-1}) \approx \frac{(X_t - X_{t-1})}{X_{t-1}} = \text{cambio relativo en } X$$

## 6A.4 Fórmulas para calcular la tasa de crecimiento

Sea la variable $Y$ una función del tiempo, $Y = \mathrm{f}(t)$, donde $t$ denota tiempo. La tasa de crecimiento instantánea (es decir, en un momento dado) de $Y$, $g_Y$ se define como

$$g_Y = \frac{dY/dt}{Y} = \frac{1}{Y}\frac{dY}{dt} \tag{29}$$

Observe que si multiplicamos $g_Y$ por 100, obtenemos la tasa porcentual de crecimiento, donde $\frac{dY}{dt}$ es la tasa de cambio de $Y$ respecto del tiempo.

Ahora bien, si $\ln Y = \ln \mathrm{f}(t)$, donde ln representa el logaritmo natural, entonces

$$\frac{d \ln Y}{dt} = \frac{1}{Y}\frac{dY}{dt} \tag{30}$$

Esto es lo mismo que la ecuación (29).

Por tanto, las transformaciones logarítmicas son muy útiles para calcular tasas de crecimiento, en especial si $Y$ es una función de otras variables dependientes del tiempo, como demostrará el siguiente ejemplo. Sea

$$Y = X \cdot Z \tag{31}$$

donde $Y$ es el PIB nominal, $X$ el PIB real y $Z$ el factor de deflación de los precios (PIB). En otras palabras, el PIB nominal es el PIB real multiplicado por el factor de deflación de los precios (PIB). Todas estas variables son funciones del tiempo, pues varían con su transcurso.

Ahora, al tomar los logaritmos en ambos miembros de la ecuación (31), obtenemos:

$$\ln Y = \ln X + \ln Z \tag{32}$$

Diferenciamos la ecuación (32) respecto del tiempo y nos da

$$\frac{1}{Y}\frac{dY}{dt} = \frac{1}{X}\frac{dX}{dt} + \frac{1}{Z}\frac{dZ}{dt} \tag{33}$$

esto es, $g_Y = g_X + g_Z$, donde $g$ representa la tasa de crecimiento.

De manera verbal, la tasa de crecimiento instantánea de $Y$ es igual a la suma de la tasa de crecimiento instantánea de $X$ más la tasa de crecimiento instantánea de $Z$. En el presente ejemplo, la tasa de crecimiento instantánea del PIB nominal es igual a la suma de la tasa de crecimiento instantánea del PIB real y la tasa de crecimiento instantánea del factor de deflación de los precios del PIB.

En términos más generales, la tasa de crecimiento instantánea de un producto es la suma de las tasas de crecimiento instantáneas de sus componentes. Esto puede generalizarse al producto de más de dos variables.

Asimismo, si tenemos

$$Y = \frac{X}{Z} \tag{34}$$

$$\frac{1}{Y}\frac{dY}{dt} = \frac{1}{X}\frac{dX}{dt} - \frac{1}{Z}\frac{dZ}{dt} \tag{35}$$

esto es, $g_Y = g_X - g_Z$. En otras palabras, la tasa instantánea de crecimiento de $Y$ es la diferencia entre la tasa de crecimiento instantánea de $X$ menos la tasa de crecimiento instantánea de $Z$. Por tanto, si $Y$ = ingreso *per cápita*, $X$ = PIB y $Z$ = población, la tasa de crecimiento instantánea del ingreso *per cápita* es igual a la tasa de crecimiento del PIB menos la tasa de crecimiento instantánea de la población.

Ahora, sea $Y = X + Z$. ¿Qué tasa de crecimiento tiene $Y$? Sea $Y$ = empleo total, $X$ = obreros empleados y $Z$ = empleados administrativos. Como

$$\ln(X + Z) \neq \ln X + \ln Y,$$

no es fácil calcular la tasa de crecimiento de $Y$, pero con algunas operaciones algebraicas se demuestra que

$$g_Y = \frac{X}{X+Z} g_X + \frac{Z}{X+Z} g_Z \tag{36}$$

Esto es, la tasa de crecimiento de una suma es un promedio ponderado de las tasas de crecimiento de sus componentes. En este ejemplo, la tasa de crecimiento del empleo total es un promedio ponderado de las tasas de crecimiento del empleo de obreros y empleados administrativos, y sus respectivas ponderaciones equivalen a la parte que representa cada componente del empleo total.

## 6A.5 Modelo de regresión Box-Cox

Considere el siguiente modelo de regresión:

$$Y_i^\lambda = \beta_1 + \beta_2 X_i + u_i \quad Y > 0 \tag{37}$$

donde λ (la letra griega lambda) es un parámetro que puede ser negativo, cero o positivo. Como $Y$ está elevado a la potencia λ, obtendremos varias transformaciones de $Y$, según el valor de λ.

La ecuación (37) se conoce como modelo de regresión Box-Cox, en honor de los estadísticos Box y Cox.[1] Según el valor de λ, tenemos los siguientes modelos de regresión, que se muestran en forma tabular:

| Valor de λ | Modelo de regresión |
|---|---|
| 1 | $Y_i = \beta_1 + \beta_2 X_i + u_i$ |
| 2 | $Y_i^2 = \beta_1 + \beta_2 X_i + u_i$ |
| 0.5 | $\sqrt{Y_i} = \beta_1 + \beta_2 X_i + u_i$ |
| 0 | $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$ |
| −0.5 | $\frac{1}{\sqrt{Y_i}} = \beta_1 + \beta_2 X_i + u_i$ |
| −1.0 | $\frac{1}{Y_i} = \beta_1 + \beta_2 X_i + u_i$ |

Como se aprecia, los modelos lineal y log-lineal son casos especiales de la familia de transformaciones Box-Cox.

Por supuesto, tales transformaciones son aplicables a las variables $X$. Es interesante notar que cuando λ es igual a cero, obtenemos la transformación logarítmica de $Y$. La prueba de esto es compleja y es mejor dejarla para las referencias. (Los lectores interesados en cálculo tendrán que recordar la regla de l'Hôpital [del hospital].)

Sin embargo, ¿cómo determinamos en realidad el valor correspondiente a λ en una situación dada? No podemos estimar la ecuación (37) de forma directa, pues esto abarca no sólo los parámetros de regresión $\beta_1$ y $\beta_2$, sino también λ, que entra de manera no lineal. No obstante, se puede demostrar que con el método de máxima verosimilitud es posible estimar todos estos parámetros. Existen paquetes de regresión que calculan precisamente esto.

No insistiremos más en este tema porque el procedimiento es un tanto complejo.

Sin embargo, podemos proceder por el método de ensayo y error. Escoja varios valores de λ, transforme $Y$ en consecuencia, ejecute la regresión (37) y obtenga la suma de cuadrados de residuos (SCR) de cada regresión transformada. Elija el valor de λ que dé la SCR mínima.[2]

[1] G.E.P. Box y D.R. Cox, "An Analysis of Transformations", *Journal of the Royal Statistical Society*, B26, 1964, pp. 211-243.

[2] Para un análisis accesible, consulte John Neter, Michael Kutner, Christopher Nachtsheim y William Wasserman, *Applied Linear Regression Models*, 3a. ed., Richard D. Irwin, Chicago, 1996.

Capítulo 7

# Análisis de regresión múltiple: el problema de estimación

El modelo con dos variables, estudiado con amplitud en los capítulos anteriores, suele ser inadecuado en la práctica. Es el caso del ejemplo consumo-ingreso (ejemplo 3.1), donde se supuso implícitamente que sólo el ingreso $X$ se relaciona con el consumo $Y$. Pero la teoría económica rara vez es tan simple, pues, además del ingreso, muchas otras variables probablemente afectan el gasto de consumo. Un ejemplo obvio es la riqueza del consumidor. Para citar otro ejemplo, es probable que la demanda de un bien dependa no sólo de su propio precio sino también de los precios de otros bienes competitivos o complementarios, del ingreso del consumidor, de la condición social, etc. Por consiguiente, se necesita ampliar el modelo simple de regresión con dos variables para considerar modelos con más de dos variables. La adición de variables conduce al análisis de los modelos de regresión múltiple, es decir, modelos en los cuales la variable dependiente, o regresada, $Y$, depende de dos o más variables explicativas, o regresoras.

El modelo de regresión múltiple más sencillo posible es la regresión de tres variables, con una variable dependiente y dos variables explicativas. En este capítulo y en el siguiente estudiaremos este modelo. Durante todo el análisis, trataremos con modelos de regresión lineal múltiple, es decir, modelos lineales en los parámetros, que pueden ser o no lineales en las variables.

## 7.1 Modelo con tres variables: notación y supuestos

Al generalizar la función de regresión poblacional (FRP) de dos variables (2.4.2), podemos escribir la FRP de tres variables así:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{7.1.1}$$

donde $Y$ es la variable dependiente, $X_2$ y $X_3$ las variables explicativas (o regresoras), $u$ es el término de perturbación estocástica, e $i$ la $i$-ésima observación; en caso de que los datos sean series de tiempo, el subíndice $t$ denotará la $t$-ésima observación.[1]

[1] Para efectos de simetría notacional, la ecuación (7.1.1) se escribe también como

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

donde $X_{1i} = 1$ para todo $i$.

En la ecuación (7.1.1), $\beta_1$ es el término del intercepto. Como es usual, este término da el efecto medio o promedio sobre $Y$ de todas las variables excluidas del modelo, aunque su interpretación mecánica sea el valor promedio de $Y$ cuando $X_2$ y $X_3$ se igualan a cero. Los coeficientes $\beta_2$ y $\beta_3$ se denominan **coeficientes de regresión parcial,** y su significado se explicará en breve.

Continuamos operando dentro del marco del modelo clásico de regresión lineal (MCRL), presentado en el capítulo 3. Específicamente, suponemos lo siguiente:

**SUPUESTOS**

1. Modelo de regresión lineal, o *lineal en los parámetros.* **(7.1.2)**
2. Valores fijos de $X$ o valores de $X$ independientes del término de error. En este caso, esto significa que se requiere covarianza cero entre $u_i$ y cada variable $X$.

$$\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = 0 \qquad \textbf{(7.1.3)}^2$$

3. Valor medio de la perturbación $u_i$ igual a cero.

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \qquad \text{por cada } i \qquad \textbf{(7.1.4)}$$

4. Homoscedasticidad o varianza constante de $u_i$.

$$\text{var}(u_i) = \sigma^2 \qquad \textbf{(7.1.5)}$$

5. No autocorrelación, o correlación serial, entre las perturbaciones.

$$\text{cov}(u_i, u_j) = 0 \qquad i \neq j \qquad \textbf{(7.1.6)}$$

6. El número de observaciones $n$ debe ser mayor que el de parámetros por estimar, que en el presente caso son 3. **(7.1.7)**
7. Debe haber variación en los valores de las variables $X$. **(7.1.8)**

También abordaremos otros dos requisitos.

8. No debe haber colinealidad exacta entre las variables $X$.

No hay **relación lineal exacta** entre $X_2$ y $X_3$ **(7.1.9)**

En la sección 7.7 dedicaremos más tiempo a analizar el supuesto final.

9. No hay *sesgo de especificación.*

El modelo está especificado correctamente. **(7.1.10)**

El fundamento de los supuestos (7.1.2) a (7.1.10) es el mismo que se explicó en la sección 3.2. El supuesto (7.1.9), que establece la no existencia de una relación lineal exacta entre $X_2$ y $X_3$, se conoce técnicamente como supuesto de *no colinealidad*, o de **no multicolinealidad** cuando hay más de una relación lineal exacta.

Informalmente, la no colinealidad significa que ninguna de las regresoras puede escribirse como combinación lineal *exacta* de las regresoras restantes en el modelo.

De manera formal, la no colinealidad significa que no existe un conjunto de números $\lambda_2$ y $\lambda_3$, al menos uno diferente de cero, tales que

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \qquad \textbf{(7.1.11)}$$

[2] Este supuesto se cumple automáticamente si $X_2$ y $X_3$ son no estocásticas y (7.1.4) se mantiene.

Si hay dicha relación lineal, se dice que $X_2$ y $X_3$ son **colineales** o linealmente dependientes. Por otra parte, si (7.1.11) se cumple sólo cuando $\lambda_2 = \lambda_3 = 0$, se dice que $X_2$ y $X_3$, son *linealmente independientes*.

Así, si

$$X_{2i} = -4X_{3i} \qquad \text{o} \qquad X_{2i} + 4X_{3i} = 0 \tag{7.1.12}$$

las dos variables son linealmente dependientes, y si se incluyen ambas en un modelo de regresión, tendremos colinealidad perfecta o una relación lineal exacta entre las dos regresoras.

Aunque consideraremos con más detalle el problema de multicolinealidad en el capítulo 10, es fácil captar intuitivamente la lógica del supuesto de no multicolinealidad. Suponga que en (7.1.1) $Y$, $X_2$ y $X_3$ representan el gasto de consumo, el ingreso y la riqueza del consumidor, respectivamente. Al postular que el gasto de consumo está relacionado linealmente con el ingreso y la riqueza, la teoría económica supone que los dos anteriores pueden tener alguna influencia independiente sobre el consumo. De no ser así, no tiene sentido incluir ambas variables, ingreso y riqueza, en el modelo. En la situación extrema, si existe una relación lineal exacta entre ingreso y riqueza, sólo hay una variable independiente, no dos, y no hay forma de evaluar la influencia *separada* del ingreso y de la riqueza sobre el consumo. Para ver esto claramente, sea $X_{3i} = 2X_{2i}$ en la regresión consumo-ingreso-riqueza. Entonces, la regresión (7.1.1) resulta ser

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3)X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned} \tag{7.1.13}$$

donde $\alpha = (\beta_2 + 2\beta_3)$. Es decir, de hecho se tiene una regresión de dos variables y no de tres. Además, si se hace la regresión (7.1.13) y se obtiene $\alpha$, no hay forma de estimar la influencia separada de $X_2$ ($= \beta_2$) y $X_3$ ($= \beta_3$) sobre $Y$, pues $\alpha$ da la *influencia combinada* de $X_2$ y $X_3$ sobre $Y$.[3]

En resumen, el supuesto de no multicolinealidad requiere que en la FRP se incluyan solamente las variables que no sean funciones lineales exactas de alguna variable del modelo. Aunque analizaremos este tema con más detalle en el capítulo 10, vale la pena notar un par de puntos aquí:

Primero, el supuesto de que no hay multicolinealidad pertenece al modelo teórico (es decir, a la FRP). En la práctica, cuando se recopilan datos para el análisis empírico, no hay una garantía de que no existirán correlaciones entre las regresoras. De hecho, en la mayor parte del trabajo aplicado casi es imposible encontrar dos o más variables (económicas) que quizá no estén correlacionadas en alguna medida, como mostraremos en los ejemplos ilustrativos más adelante en este capítulo. Lo que se requiere es que no haya una relación exacta entre las regresoras, como en la ecuación (7.1.12).

En segundo lugar, tenga presente que sólo hablamos de relaciones lineales perfectas entre dos o más variables. La multicolinealidad no es válida para relaciones no lineales entre variables. Suponga que $X_{3i} = X_{2i}^2$. Esto no viola el supuesto de no colinealidad perfecta, en vista de que la relación entre las variables es no lineal.

[3] En términos matemáticos, $\alpha = (\beta_2 + 2\beta_3)$ es una ecuación con dos incógnitas y no hay una forma *única* de estimar $\beta_2$ y $\beta_3$ a partir de la $\alpha$ estimada.

## 7.2 Interpretación de la ecuación de regresión múltiple

Con los supuestos del modelo de regresión clásico, se cumple que, al tomar la esperanza condicional de $Y$ en ambos lados de (7.1.1), obtenemos

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} \tag{7.2.1}$$

Expresado en palabras, de (7.2.1) se obtiene la **media condicional o el valor esperado de $Y$ condicionado a los valores dados o fijos de las variables** $X_2$ y $X_3$. Por consiguiente, como en el caso de dos variables, el análisis de regresión múltiple es el análisis de regresión condicional sobre los valores fijos de las variables explicativas, y lo que obtenemos es el valor promedio o la media de $Y$, o la respuesta media de $Y$ a los valores dados de las regresoras $X$.

## 7.3 Significado de los coeficientes de regresión parcial

Como ya mencionamos, los coeficientes de regresión $\beta_2$ y $\beta_3$ se conocen como **coeficientes de regresión parcial** o **coeficientes parciales de pendiente**. El significado del coeficiente de regresión parcial es el siguiente: $\beta_2$ mide el *cambio* en el valor de la media de $Y$, $E(Y)$, por unidad de cambio en $X_2$, con $X_3$ constante. Expresado de otra forma, proporciona el efecto "directo" o "neto" que tiene una unidad de cambio de $X_2$ sobre el valor medio de $Y$, neto de cualquier efecto que $X_3$ pueda ejercer en la media $Y$. De igual forma, $\beta_3$ mide el cambio en el valor medio de $Y$ por unidad de cambio en $X_3$, cuando el valor de $X_2$ se conserva constante.[4] Es decir, da el efecto "directo" o "neto" de una unidad de cambio en $X_3$ sobre el valor medio de $Y$, neto de cualquier efecto que $X_2$ pudiera tener sobre la media $Y$.[5]

En realidad, ¿cómo haríamos para conservar constante la influencia de una regresora? Para averiguarlo, retomemos el ejemplo de la mortalidad infantil. Recuerde que en ese ejemplo $Y =$ mortalidad infantil (MI), $X_2 =$ PIB *per cápita* (PIBPC) y $X_3 =$ tasa de alfabetización en las mujeres (TAM). Suponga que deseamos conservar constante la influencia de la TAM. Como TAM puede tener algún efecto sobre MI así como sobre PIBPC para cualesquiera datos concretos, lo que podemos hacer es eliminar la influencia (lineal) que TAM ejerce tanto en MI como en PIBPC, al hacer la regresión de MI sobre TAM y la regresión de PIBPC sobre TAM en forma separada, y luego buscar los residuos obtenidos de tales regresiones. Con los datos proporcionados en la tabla 6.4, obtenemos las siguientes regresiones:

$$\begin{aligned} \widehat{\text{MI}}_i &= 263.8635 \;-\; 2.3905\ \text{TAM}_i + \hat{u}_{1i} \\ \text{ee} &= (12.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6695 \end{aligned} \tag{7.3.1}$$

donde $\hat{u}_{1i}$ representa el término residual de esta regresión.

$$\begin{aligned} \widehat{\text{PIBPC}}_i &= -39.3033 \;+\; 28.1427\ \text{TAM}_i + \hat{u}_{2i} \\ \text{ee} &= (734.9526) \quad (12.8211) \qquad r^2 = 0.0721 \end{aligned} \tag{7.3.2}$$

donde $\hat{u}_{2i}$ representa el término residual de esta regresión.

[4] El lector con conocimientos de cálculo se dará cuenta de inmediato de que $\beta_2$ y $\beta_3$ son las derivadas parciales de $E(Y \mid X_2, X_3)$ respecto de $X_2$ y $X_3$.

[5] A propósito, las expresiones *mantener constante, estar controlado por, permitir o tomar en cuenta la influencia de, corregir la influencia de* y *eliminar la influencia de* son sinónimos y se utilizarán de manera indistinta a lo largo de esta obra.

Ahora

$$\hat{u}_{1i} = (\text{MI}_i - 263.8635 + 2.3905\ \text{TAM}_i) \tag{7.3.3}$$

representa la parte de MI que queda después de anular la influencia (lineal) de TAM. De igual forma,

$$\hat{u}_{2i} = (\text{PIBPC}_i + 39.3033 - 28.1427\ \text{TAM}_i) \tag{7.3.4}$$

representa la parte del PIBPC que queda después de eliminar la influencia (lineal) de TAM.

Por consiguiente, si ahora se hace la regresión de $\hat{u}_{1i}$ o $\hat{u}_{2i}$, que están "purificadas" de la influencia (lineal) de TAM, ¿no obtendríamos el efecto neto del PIBPC sobre MI? De hecho, así sucede (véase el apéndice 7A, sección 7A.2). Los resultados de la regresión son los siguientes:

$$\begin{aligned} \hat{\hat{u}}_{1i} &= -0.0056\hat{u}_{2i} \\ \text{ee} &= \ (0.0019) \qquad r^2 = 0.1152 \end{aligned} \tag{7.3.5}$$

*Nota:* Esta regresión no tiene el término de intercepto porque el valor medio de los residuos MCO $\hat{u}_{1i}$ y $\hat{u}_{2i}$ son cero (¿por qué?).

El coeficiente de la pendiente de $-0.0056$ ahora proporciona el efecto "verdadero" o neto de una unidad de cambio en PIBPC sobre MI, o la verdadera pendiente de MI respecto de PIBPC. Es decir, proporciona el coeficiente de regresión parcial de MI respecto de PIBPC, $\beta_2$.

Quienes deseen determinar los coeficientes de la regresión parcial de MI respecto de TAM pueden repetir el procedimiento anterior: primero deben efectuar la regresión de MI sobre PIBPC para obtener los residuos de esta regresión ($\hat{u}_{1i}$), después, la regresión de TAM sobre PIBPC para obtener los residuos de esta regresión ($\hat{u}_{2i}$), y por último, la regresión $\hat{u}_{1i}$ sobre $\hat{u}_{2i}$. Estoy seguro de que los lectores comprendieron la idea.

¿Tenemos que llevar a cabo este procedimiento de múltiples pasos siempre que deseemos conocer el verdadero coeficiente de la regresión parcial? Por fortuna, no, pues el mismo resultado se obtiene de forma muy rápida y rutinaria mediante el procedimiento MCO analizado en esta sección. El método de varios pasos que acabamos de esbozar tiene sólo fines pedagógicos, a fin de exponer el significado del coeficiente de regresión "parcial".

## 7.4 Estimación de MCO y MV de los coeficientes de regresión parcial

Para estimar los parámetros del modelo de regresión con tres variables (7.1.1) consideraremos primero el método de mínimos cuadrados ordinarios (MCO), presentado en el capítulo 3, y luego, brevemente, el método de máxima verosimilitud (MV), estudiado en el capítulo 4.

### Estimadores de MCO

Para encontrar los estimadores de MCO, escribamos primero la función de regresión muestral (FRM) correspondiente a la FRP de (7.1.1) de la siguiente manera:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{7.4.1}$$

donde $\hat{u}_i$ es el término residual, la contraparte muestral del término de perturbación estocástico $u_i$.

Como se vio en el capítulo 3, el procedimiento MCO consiste en seleccionar los valores desconocidos de los parámetros de forma que la suma de cuadrados de los residuos (SCR) $\sum \hat{u}_i^2$ sea lo más pequeña posible. Simbólicamente,

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \tag{7.4.2}$$

donde la expresión para la SCR se obtiene por simple manipulación algebraica de (7.4.1).

El procedimiento más directo para obtener los estimadores que reducen (7.4.2) es diferenciarla respecto de las incógnitas, igualar a cero las expresiones resultantes y resolverlas al mismo tiempo. Como se muestra en el apéndice 7A, sección 7A.1, de este procedimiento se obtienen las siguientes *ecuaciones normales* [comparables con las ecuaciones (3.1.4) y (3.1.5)]:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 \tag{7.4.3}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \tag{7.4.4}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \tag{7.4.5}$$

De la ecuación (7.4.3) vemos al instante que

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \tag{7.4.6}$$

que es el estimador de MCO del intercepto poblacional $\beta_1$.

Conforme a la convención de permitir que las letras minúsculas denoten desviaciones de las medias muestrales, se derivan las siguientes fórmulas de las ecuaciones normales (7.4.3) a (7.4.5):

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.7}$$ [6]

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.8}$$

que dan los estimadores de MCO de los coeficientes de regresión parcial poblacionales, $\beta_2$ y $\beta_3$, respectivamente.

A propósito, observe lo siguiente: 1) Las ecuaciones (7.4.7) y (7.4.8) son simétricas por naturaleza, porque una se obtiene de la otra mediante el cambio de papeles de $X_2$ y $X_3$; 2) los denominadores en estas dos ecuaciones son idénticos; y 3) el caso de tres variables es una extensión natural del caso de dos variables.

[6] Este estimador es igual al de (7.3.5), como se muestra en el apéndice 7A, sección 7A.2.

## Varianzas y errores estándar de los estimadores de MCO

Después de obtener los estimadores de MCO de los coeficientes de regresión parcial derivamos las varianzas y los errores estándar de los estimadores como se indica en el apéndice 3A.3. Igual que en el caso de dos variables, se necesitan los errores estándar para dos fines principales: establecer intervalos de confianza y probar hipótesis estadísticas. Las fórmulas pertinentes son las siguientes:[7]

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i}x_{3i}\right)^2}\right] \cdot \sigma^2 \tag{7.4.9}$$

$$\operatorname{ee}(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)} \tag{7.4.10}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \tag{7.4.11}$$

o, en forma equivalente,

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \tag{7.4.12}$$

donde $r_{23}$ es el coeficiente de correlación muestral entre $X_2$ y $X_3$, como se define en el capítulo 3.[8]

$$\operatorname{ee}(\hat{\beta}_2) = +\sqrt{\operatorname{var}(\hat{\beta}_2)} \tag{7.4.13}$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \tag{7.4.14}$$

o, en forma equivalente,

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2\left(1 - r_{23}^2\right)} \tag{7.4.15}$$

$$\operatorname{ee}(\hat{\beta}_3) = +\sqrt{\operatorname{var}(\hat{\beta}_3)} \tag{7.4.16}$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}} \tag{7.4.17}$$

En todas estas fórmulas, $\sigma^2$ es la varianza (homoscedástica) de las perturbaciones poblacionales $u_i$.

Según el argumento del apéndice 3A, sección 3A.5, el lector puede verificar que un estimador insesgado de $\sigma^2$ está dado por

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3} \tag{7.4.18}$$

[7] Las derivaciones de estas fórmulas son más sencillas con notación matricial. Los lectores con mayores conocimientos pueden consultar el **apéndice C**.

[8] Con la definición de *r* dada en el capítulo 3, tenemos

$$r_{23}^2 = \frac{\left(\sum x_{2i}x_{3i}\right)^2}{\sum x_{2i}^2 \sum x_{3i}^2}$$

Observe la similitud entre este estimador de $\sigma^2$ y su correspondiente en el caso de dos variables $[\hat{\sigma}^2 = (\sum \hat{u}_i^2)/(n-2)]$. Los grados de libertad son ahora $(n-3)$ porque, para calcular $\sum \hat{u}_i^2$, se deben estimar primero $\beta_1$, $\beta_2$ y $\beta_3$, los cuales consumen 3 gl. (El argumento es muy general. Así, en el caso de cuatro variables, los gl serán $n-4$.)

El estimador $\hat{\sigma}^2$ se calcula de (7.4.18) una vez que se dispone de los residuos, pero también se obtiene, más rápido, con la siguiente relación (para una prueba, véase el apéndice 7A, sección 7A.3):

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

que es la contraparte de tres variables de la relación dada en (3.3.6).

## Propiedades de los estimadores de MCO

Las propiedades de los estimadores de MCO del modelo de regresión múltiple son similares a las del modelo con dos variables. Específicamente:

1. La línea (superficie) de regresión de tres variables pasa a través de las medias de $\bar{Y}$, $\bar{X}_2$ y $\bar{X}_3$, lo cual se hace evidente en (7.4.3) [compárese con la ecuación (3.1.7) del modelo con dos variables]. Esta propiedad generalmente se mantiene. Así, en el modelo de regresión lineal con $k$ variables [una regresada y $(k-1)$ regresoras],

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{7.4.20}$$

se tiene que

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \hat{X}_3 - \cdots - \beta_k \bar{X}_k \tag{7.4.21}$$

2. El valor medio de $Y_i$ estimado $Y_i$ $(= \hat{Y}_i)$ es igual al valor medio de $Y_i$ observado, lo cual es fácil de demostrar:

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \qquad \text{(¿Por qué?)} \\ &= \bar{Y} + \hat{\beta}_2 (X_{2i} - \bar{X}_2) + \hat{\beta}_3 (X_{3i} - \bar{X}_3) \\ &= \bar{Y} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \end{aligned} \tag{7.4.22}$$

donde, como es usual, las letras minúsculas indican los valores de las variables expresadas como desviaciones de sus medias respectivas.

Al sumar en ambos lados de (7.4.22) sobre los valores muestrales y dividir entre el tamaño de la muestra $n$, se tiene $\bar{\hat{Y}} = \bar{Y}$. (*Nota:* $\sum x_{2i} = \sum x_{3i} = 0$. ¿Por qué?) Observe que, en virtud de (7.4.22), podemos escribir

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \tag{7.4.23}$$

donde $\hat{y}_i = (\hat{Y}_i - \bar{Y})$.

Por consiguiente, la FRM (7.4.1) se expresa en *forma de desviaciones* como

$$y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{7.4.24}$$

3. $\sum \hat{u}_i = \bar{\hat{u}} = 0$, lo cual se verifica de (7.4.24). [*Sugerencia:* Sume ambos lados de (7.4.24) sobre los valores muestrales.]

4. Los residuos $\hat{u}_i$ no están correlacionados con $X_{2i}$ y $X_{3i}$, es decir, $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$ (véase la prueba en el apéndice 7A.1).

5. Los residuos $\hat{u}_i$ no están correlacionados con $\hat{Y}_i$; es decir, $\sum \hat{u}_i \hat{Y}_i = 0$. ¿Por qué? [*Sugerencia:* Multiplique ambos lados de (7.4.23) por $\hat{u}_i$ y sume sobre los valores muestrales.]

6. De (7.4.12) y (7.4.15) es evidente que, a medida que $r_{23}$, el coeficiente de correlación entre $X_2$ y $X_3$, se acerca a 1, las varianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$ aumentan para los valores dados de $\sigma^2$ y $\sum x_{2i}^2$ o $\sum x_{3i}^2$. En el límite, cuando $r_{23} = 1$ (es decir, la colinealidad perfecta), estas varianzas se hacen infinitas. En el capítulo 10 analizaremos a fondo las implicaciones de esto, pero ya el lector puede intuir que, a medida que aumenta $r_{23}$, resulta cada vez más difícil conocer los valores verdaderos de $\beta_2$ y $\beta_3$. [Veremos más detalles en el siguiente capítulo, pero mientras, consulte la ecuación (7.1.13).]

7. También es claro, de (7.4.12) y (7.4.15), que, para valores dados de $r_{23}$ y $\sum x_{2i}^2$ o $\sum x_{3i}^2$, las varianzas de los estimadores de MCO son directamente proporcionales a $\sigma^2$; es decir, aumentan a medida que lo hace $\sigma^2$. En forma similar, para valores dados de $\sigma^2$ y $r_{23}$, la varianza de $\hat{\beta}_2$ es inversamente proporcional a $\sum x_{2i}^2$; es decir, entre mayor sea la variación de los valores muestrales de $X_2$, menor será la varianza de $\hat{\beta}_2$ y, por consiguiente, $\beta_2$ se estima en forma más precisa. Una afirmación similar vale respecto de la varianza de $\hat{\beta}_3$.

8. Con los supuestos del modelo clásico de regresión lineal enunciados en la sección 7.1 se demuestra que los estimadores de MCO de los coeficientes de regresión parcial no solamente son lineales e insesgados, sino que también tienen varianza mínima dentro de la clase de todos los estimadores lineales insesgados. En resumen, *son MELI*. Dicho de otra forma, satisfacen el teorema de Gauss-Markov. (La prueba es similar al caso de dos variables demostrado en el apéndice 3A, sección 3A.6, y se presentará en forma más compacta con notación matricial en el **apéndice C**).

### Estimadores de máxima verosimilitud

En el capítulo 4 observamos que, según el supuesto de que las perturbaciones poblacionales, $u_i$, estén normalmente distribuidas con media cero y varianza $\sigma^2$ constante, los estimadores de máxima verosimilitud (MV) y los estimadores de MCO de los coeficientes de regresión del modelo con dos variables son idénticos. Esta igualdad se extiende a modelos con cualquier número de variables. (Para una demostración, véase el apéndice 7A, sección 7A.4.) Sin embargo, esto no vale para el estimador de $\sigma^2$. Puede demostrarse que el estimador de MV de $\sigma^2$ es $\sum \hat{u}_i^2/n$ sin importar el número de variables en el modelo, mientras que el estimador de MCO de $\sigma^2$ es $\sum \hat{u}_i^2/(n-2)$ en el caso de dos variables, $\sum \hat{u}_i^2/(n-3)$ en el caso de tres variables y $\sum \hat{u}_i^2/(n-k)$ en el caso del modelo de $k$ variables (7.4.20). En resumen, el estimador de MCO de $\sigma^2$ tiene en cuenta el número de grados de libertad, mientras que el estimador MV no lo hace. Por supuesto, si $n$ es grande, los estimadores de MV y MCO de $\sigma^2$ tienden a estar cerca uno del otro. (¿Por qué?)

## 7.5 El coeficiente múltiple de determinación $R^2$ y el coeficiente múltiple de correlación $R$

En el caso de dos variables vimos que $r^2$, definido en (3.5.5), mide la bondad de ajuste de la ecuación de regresión; es decir, da la proporción o porcentaje de la variación total en la variable dependiente $Y$ explicada por la variable (única) explicativa $X$. Esta notación de $r^2$ se extiende fácilmente a los modelos de regresión con más de dos variables. Así, en el modelo de tres variables buscamos conocer la proporción de la variación en $Y$ explicada por las variables $X_2$ y $X_3$ conjuntamente. La medida que da esta información se conoce como **coeficiente de determinación múltiple,** y se denota por $R^2$; conceptualmente se asemeja a $r^2$.

Para obtener $R^2$, podemos seguir el procedimiento para obtener $r^2$ descrito en la sección 3.5. Recuerde que

$$
\begin{aligned}
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\
&= \hat{Y}_i + \hat{u}_i
\end{aligned}
\tag{7.5.1}
$$

donde $\hat{Y}_i$ es el valor estimado de $Y_i$ a partir de la línea de regresión ajustada y es un estimador de la verdadera $E(Y_i \mid X_{2i}, X_{3i})$. Al sustituir las letras mayúsculas por minúsculas para indicar desviaciones de sus medias, la ecuación (7.5.1) se escribe como

$$
\begin{aligned}
y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \\
&= \hat{y}_i + \hat{u}_i
\end{aligned}
\tag{7.5.2}
$$

Elevamos al cuadrado (7.5.2) en ambos lados y sumamos sobre los valores muestrales para obtener

$$
\begin{aligned}
\sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i \\
&= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \qquad (\text{¿Por qué?})
\end{aligned}
\tag{7.5.3}
$$

En palabras, la ecuación (7.5.3) afirma que la suma de cuadrados total (STC) es igual a la suma de cuadrados explicada (SCE) + la suma de cuadrados de residuos (SCR). Ahora, sustituimos el equivalente de $\sum \hat{u}_i^2$ dado en la ecuación (7.4.19) y obtenemos

$$
\sum y_i^2 = \sum \hat{y}_i^2 + \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}
$$

la cual, al reordenar términos, da

$$
\text{SCE} = \sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}
\tag{7.5.4}
$$

Ahora, por definición,

$$
\begin{aligned}
R^2 &= \frac{\text{SCE}}{\text{SCT}} \\
&= \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}
\end{aligned}
\tag{7.5.5}
$$
[9]

[Compare (7.5.5) con (3.5.6).]

Como las cantidades consideradas en (7.5.5) suelen calcularse de forma rutinaria, $R^2$ se calcula sin problemas. Observe que $R^2$, al igual que $r^2$, se encuentra entre 0 y 1. Si es 1, la línea de regresión ajustada explica 100% de la variación en $Y$. Por otra parte, si es 0, el modelo no explica nada de la variación en $Y$. Sin embargo, por lo general $R^2$ se encuentra entre estos dos valores extremos. Se dice que el ajuste del modelo es "mejor" entre más cerca esté $R^2$ de 1.

---

[9] Observe que $R^2$ también se calcula de la siguiente manera:

$$
R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{(n-3)\hat{\sigma}^2}{(n-1)S_Y^2}
$$

Recuerde que, en el caso de dos variables, definimos $r$ como el coeficiente de correlación e indicamos que mide el grado de asociación (lineal) entre las dos variables. El análogo de $r$ para tres o más variables es el coeficiente de **correlación múltiple**, denotado por $R$, el cual es una medida del grado de asociación entre $Y$ y todas las variables explicativas en conjunto. Aunque $r$ puede ser positivo o negativo, $R$ siempre se considera positivo. En la práctica, sin embargo, $R$ tiene poca importancia. La medida de mayor significado es $R^2$.

Antes de continuar, establezcamos la siguiente relación entre $R^2$ y la varianza de un coeficiente de regresión parcial en el modelo de regresión múltiple con $k$ variables dado en (7.4.20):

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \tag{7.5.6}$$

donde $\hat{\beta}_j$ es el coeficiente de regresión parcial de la regresora $X_j$ y $R_j^2$ es el $R^2$ en la regresión de $X_j$ sobre las $(k-2)$ regresoras restantes. [*Nota:* En el modelo de regresión con $k$ variables hay $(k-1)$ regresoras]. Aunque la utilidad de la ecuación (7.5.6) se verá en el capítulo 10, sobre multicolinealidad, observe que esta ecuación es sólo una extensión de la fórmula dada en (7.4.12) o (7.4.15) para el modelo de regresión con tres variables, una regresada y dos regresoras.

## 7.6 Un ejemplo ilustrativo

**EJEMPLO 7.1**
*Mortalidad infantil en relación con el PIB per cápita y la tasa de alfabetización de las mujeres*

En el capítulo 6 consideramos el comportamiento de la mortalidad infantil (MI) en relación con el PIB *per cápita* (PIBPC). Vimos que el PIBPC ejerce un impacto negativo sobre la MI, como era de esperarse. Ahora se presenta el alfabetismo femenino medido por la tasa de alfabetización de las mujeres (TAM). *A priori*, se espera que la TAM también ejerza un impacto negativo en la MI. Ahora, cuando se introducen ambas variables en el modelo, se requiere eliminar la influencia neta de cada regresora. Es decir, necesitamos estimar los coeficientes de regresión (parcial) de cada regresora. Por tanto, el modelo es:

$$\text{MI}_i = \beta_1 + \beta_2 \text{PIBPC}_i + \beta_3 \text{TAM}_i + u_i \tag{7.6.1}$$

Los datos necesarios se proporcionan en la tabla 6.4. Tenga en cuenta que la MI es el número de muertes de niños menores de 5 años por cada 1 000 nacidos vivos, el PIBPC es el PIB *per cápita* en 1980 y la TAM se mide en porcentaje. La muestra se realizó en 64 países.

Con el paquete estadístico *EViews6* se obtienen los siguientes resultados:

$$\begin{aligned} \widehat{\text{MI}}_i &= 263.6416 \;-\; 0.0056\ \text{PIBPC}_i - \;2.2316\ \text{TAM}_i \\ \text{ee} &= (11.5932) \quad (0.0019) \qquad (0.2099) \qquad R^2 = 0.7077 \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \bar{R}^2 = 0.6981^* \end{aligned} \tag{7.6.2}$$

donde las cifras en paréntesis son los errores estándar estimados. Antes de interpretar esta regresión, observe el coeficiente parcial de la pendiente del PIBPC, a saber, −0.0056. ¿No es precisamente el mismo que obtuvimos del procedimiento en tres pasos de la sección anterior [véase la ecuación (7.3.5)]? ¿Esto debe sorprender? Y no nada más eso, sino que también los dos errores estándar son justamente los mismos, lo cual de nuevo no debe causar sorpresa alguna. Hicimos eso mismo pero sin el engorroso procedimiento de tres pasos.

* Al respecto, véase la sección 7.8.

Ahora interpretemos estos coeficientes de regresión: −0.0056 es el coeficiente de regresión parcial del PIBPC e indica que, si se mantiene constante la influencia de la TAM, conforme el PIBPC se incrementa, por ejemplo en un dólar en promedio, la mortalidad infantil disminuye en 0.0056 unidades. Para interpretar esto desde el punto de vista económico, si el PIB *per cápita* se incrementara 1 000 dólares, en promedio, el número de muertes de niños menores de 5 años se reduciría a 5.6 por cada 1 000 nacimientos vivos. El coeficiente −2.2316 señala que si la influencia del PIBPC se mantiene constante, el número de muertes de niños menores de 5 años disminuiría, en promedio, 2.23 por cada 1 000 nacimientos vivos, si la tasa de alfabetización en las mujeres subiera un punto porcentual. El valor del intercepto de alrededor de 263, si se interpretara de una forma mecanicista, significaría que si los valores del PIBPC y de la TAM fuesen cero, la mortalidad infantil promedio sería de más o menos 263 muertes por cada 1 000 nacimientos vivos. Por supuesto, tal interpretación debe tomarse con mucho cuidado. Cualquiera puede inferir que si dos regresoras tuviesen un valor cero, la mortalidad infantil sería muy alta, lo cual tiene sentido. El valor de $R^2$ de casi 0.71 significa que casi 71% de la variación en la mortalidad infantil se explica mediante el PIBPC y la TAM, lo cual es un gran porcentaje si se considera que el valor máximo que puede tener $R^2$ es 1. De todo lo dicho hasta aquí, los resultados de la regresión tienen sentido.

¿Y qué pasa con la importancia de los coeficientes estimados? Abordaremos este tema en el capítulo 8. Como verá, dicho capítulo es una extensión del capítulo 5, que aborda el modelo de dos variables. También verá que hay algunas diferencias importantes, en cuanto a inferencia estadística (es decir, pruebas de hipótesis), entre el modelo de regresión con dos variables y el multivariado.

## Regresión sobre variables estandarizadas

En el capítulo anterior presentamos el tema de la regresión sobre variables estandarizadas y afirmamos que el análisis se podía extender a las regresiones multivariadas. Recuerde que una variable es estandarizada o se expresa en unidades de desviación estándar si se expresa en términos de desviación de su media y se dividió entre su desviación estándar.

Para el ejemplo de la mortalidad infantil, los resultados son los siguientes:

$$\widehat{\text{MI}}^* = -\ 0.2026\ \text{PIBPC}_i^* - 0.7639\ \text{TAM}_i^* \qquad \textbf{(7.6.3)}$$

$$\text{ee} = \quad (0.0713) \qquad\qquad (0.0713) \qquad r^2 = 0.7077$$

*Nota:* Las variables con asterisco están estandarizadas. También observe que no existe término de intercepto en el modelo por las razones que ya analizamos en el capítulo anterior.

Como se aprecia en esta regresión, si se conserva constante la TAM, un incremento igual a una desviación estándar en el PIBPC propicia, en promedio, una disminución de 0.2026 desviación estándar en la MI. De manera semejante, si se conserva al PIBPC constante, un incremento igual a una desviación estándar en la TAM, en promedio, causará una disminución de 0.7639 de desviación estándar en la MI. En términos relativos, el alfabetismo en las mujeres tiene un mayor impacto en la mortalidad infantil que el PIB *per cápita*. Aquí se advierte la ventaja de utilizar variables estandarizadas, pues la estandarización hace que todas las variables tengan una medida común, en vista de que todas las variables estandarizadas tienen medias iguales a cero y varianzas unitarias.

## Efecto sobre la variable dependiente de un cambio unitario en más de una regresora

Antes de continuar, suponga que deseamos averiguar qué pasaría con la tasa de mortalidad infantil si el PIBPC y la TAM se incrementaran de manera simultánea. Suponga que el PIB *per cápita* aumenta un dólar y, al mismo tiempo, la tasa de alfabetización de las mujeres aumenta un punto porcentual. ¿Cuál sería el efecto de este cambio simultáneo en la tasa de mortalidad infantil?

Para averiguarlo, sólo hay que multiplicar los coeficientes del PIBPC y la TAM por los cambios propuestos y sumar los términos resultantes. En el ejemplo, esto da:

$$-0.0056(1) - 2.2316(1) = 2.2372$$

Es decir, como resultado de este cambio simultáneo en el PIBPC y la TAM, el número de muertes de niños menores de 5 años disminuiría en casi 2.24 muertes.

En términos más generales, si deseamos conocer el efecto total sobre la variable dependiente de un cambio unitario en más de una regresora, todo lo que hay que hacer es multiplicar los coeficientes de dichas regresoras por los cambios propuestos y sumar los productos. Observe que el término del intercepto no interviene en estos cálculos. (¿Por qué?)

## 7.7 Regresión simple en el contexto de regresión múltiple: introducción al sesgo de especificación

El supuesto (7.1.10) del MCRL plantea que el modelo de regresión del análisis está "bien" especificado; es decir, no hay **sesgo o error de especificación** (véase algunos comentarios introductorios en el capítulo 3). Aunque analizaremos con más detalle el tema del análisis de especificación en el capítulo 13, el ejemplo ilustrativo de la sección anterior da una oportunidad no sólo para entender la importancia del supuesto (7.1.10), sino también para aclarar de manera adicional el significado del coeficiente de regresión parcial y presentar una introducción formal al tema del sesgo de especificación.

Suponga que (7.6.1) es el modelo "verdadero" que explica el comportamiento de la mortalidad infantil en relación con el PIB *per cápita* y la tasa de alfabetización de las mujeres (TAM). También suponga que se hace caso omiso de la TAM y que se estima la siguiente regresión simple:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \tag{7.7.1}$$

donde $Y =$ MI y $X_2 =$ PIBPC.

Como (7.6.1) es el verdadero modelo, al estimar (7.7.1) se cometería un error de especificación, el cual consiste en *omitir* la variable $X_3$, la tasa de alfabetización de las mujeres. Observe que usamos diferentes símbolos para los parámetros (las alfas) en (7.7.1) para distinguirlas de los parámetros verdaderos (las betas) dadas en (7.6.1).

Ahora, ¿$\alpha_2$ proporcionará un estimado insesgado del verdadero impacto de PIBPC, dado por $\beta_2$ en el modelo (7.6.1)? En otras palabras, ¿$E(\hat{\alpha}_2) = \beta_2$, donde $\hat{\alpha}_2$ es el valor estimado de $\alpha_2$? Es decir, ¿el coeficiente del PIBPC en (7.7.1) suministra una estimación insesgada del verdadero impacto del PIBPC sobre la MI, si sabemos que se omitió la variable $X_3$ (TAM) del modelo? Como se puede sospechar, *en general* $\hat{\alpha}_2$ no será un estimador insesgado del verdadero $\beta_2$. Para tener una ligera idea del sesgo, hagamos la regresión (7.7.1), la cual da los siguientes resultados:

$$\begin{aligned} \widehat{\text{TM}}_i &= 157.4244 \quad - \quad 0.0114\ \text{PIBPC}_i \\ \text{ee} &= \quad (9.8455) \qquad (0.0032) \qquad r^2 = 0.1662 \end{aligned} \tag{7.7.2}$$

Observe varias cosas respecto de esta regresión en comparación con la regresión múltiple "verdadera" (7.6.1):

1. En términos absolutos (es decir, omitiendo el signo), el coeficiente del PIBPC se incrementó de 0.0056 a 0.0114, casi el doble.

2. Los errores estándar son diferentes.
3. Los valores del intercepto son distintos.
4. Los valores $r^2$ son muy distintos, aunque, por lo general, conforme aumenta el número de regresoras en el modelo, se incrementan los valores $r^2$.

Ahora suponga que se hace la regresión de la mortalidad infantil respecto de la tasa de alfabetismo en las mujeres sin tener en cuenta la influencia del PIBPC. Se obtiene lo siguiente:

$$\begin{aligned} \widehat{\text{TM}}_i &= 263.8635 \;-\; 2.3905\ \text{TAM}_i \\ \text{ee} &= (21.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6696 \end{aligned} \tag{7.7.3}$$

De nuevo, si compara los resultados de esta regresión (mal especificada) con la regresión múltiple "verdadera", verá que son distintos, aunque las diferencias en este caso no resultan tan notables como en la regresión (7.7.2).

*Lo importante es estar consciente de que puede haber serias consecuencias si se elige mal un modelo*. Analizaremos con más detalle este punto en el capítulo 13, sobre los errores de especificación.

## 7.8 $R^2$ y $R^2$ ajustada

Una propiedad importante de $R^2$ es que es una función no decreciente del número de variables explicativas o de regresoras presentes en el modelo; a medida que aumenta el número de regresoras, $R^2$ aumenta casi invariablemente y nunca disminuye. Planteado de otra forma, una variable adicional $X$ no reduce $R^2$. Compare, por ejemplo, la regresión (7.7.2) o la (7.7.3) con la (7.6.2). Para ver esto, recuerde la definición del coeficiente de determinación:

$$\begin{aligned} R^2 &= \frac{\text{SCE}}{\text{SCT}} \\ &= 1 - \frac{\text{SCR}}{\text{SCT}} \\ &= 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \end{aligned} \tag{7.8.1}$$

Ahora, $\sum y_i^2$ es independiente del número de variables $X$ en el modelo porque es sólo $\sum (Y_i - \bar{Y})^2$. SCR, $\sum \hat{u}_i^2$, sin embargo, depende del número de regresoras presentes en el modelo. Por intuición, es claro que, a medida que aumenta el número de variables $X$, es más probable que disminuya $\sum \hat{u}_i^2$ (al menos, que no aumente); por tanto, $R^2$, como se definió en (7.8.1), aumenta. En vista de esto, al comparar dos modelos de regresión con la *misma variable dependiente* pero un número diferente de variables $X$, se debe tener mucho cuidado al escoger el modelo con la $R^2$ más alta.

Para comparar dos términos $R^2$ se debe tener en cuenta el número de variables $X$ presentes en el modelo. Esto se verifica con facilidad si consideramos un coeficiente de determinación alterno, que es el siguiente:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)} \tag{7.8.2}$$

donde $k =$ el número de parámetros en el modelo *incluyendo el término de intercepto*. (En la regresión con tres variables, $k = 3$. ¿Por qué?) $R^2$ definida así se conoce como **$R^2$ ajustada**, designada por $\bar{R}^2$. El término *ajustado* significa ajustado por los gl asociados a las sumas de cuadrados que se consideran en (7.8.1): $\sum \hat{u}_i^2$ tiene $n - k$ gl en un modelo con $k$ parámetros, el cual incluye el término del intercepto y $\sum \hat{u}_i^2$ tiene $n - 1$ gl. (¿Por qué?) Para el caso de tres variables, sabemos que $\sum \hat{u}_i^2$ tiene $n - 3$ gl.

La ecuación (7.8.2) también se escribe como

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \tag{7.8.3}$$

donde $\hat{\sigma}^2$ es la varianza residual, un estimador insesgado de la verdadera $\sigma^2$, y $S_Y^2$ es la varianza muestral de $Y$.

Es fácil ver que el $\bar{R}^2$ y el $R^2$ están relacionados porque, al incluir (7.8.1) en (7.8.2), obtenemos

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{7.8.4}$$

Por la ecuación (7.8.4) de inmediato se comprende que 1) para $k > 1$, $\bar{R}^2 < R^2$, lo cual implica que, a medida que aumenta el número de variables $X$, $R^2$ ajustada aumenta menos que $R^2$ no ajustada; y que 2) $\bar{R}^2$ puede ser negativa, aunque $R^2$ es necesariamente no negativa.[10] En caso de que $\bar{R}^2$ resulte ser negativa en una aplicación, su valor se toma como cero.

¿Cuál $R^2$ debe utilizarse en la práctica? Como anota Theil:

> . . . es una buena costumbre utilizar $\bar{R}^2$ en lugar de $R^2$ porque $R^2$ tiende a dar una imagen demasiado optimista del ajuste de la regresión, en particular cuando el número de variables explicativas no es muy pequeño comparado con el de observaciones.[11]

Pero no todos comparten la opinión de Theil, pues no ofrece una justificación teórica general para la "superioridad" de $\bar{R}^2$. Por ejemplo, Goldberger argumenta que la siguiente $R^2$, denominada $R^2$ **modificada**, servirá igual.[12]

$$R^2 \text{ modificada} = (1 - k/n)R^2 \tag{7.8.5}$$

Su sugerencia es informar sobre $R^2$, $n$ y $k$, y dejar que el lector elija la forma de ajustar $R^2$ considerando $n$ y $k$.

[10] Observe, sin embargo, que si $R^2 = 1$, $\bar{R}^2 = R^2 = 1$. Cuando $R^2 = 0$, $\bar{R}^2 = (1 - k)/(n - k)$, en cuyo caso $\bar{R}^2$ es negativa si $k > 1$.

[11] Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, p. 135.

[12] Arthur S. Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991, p. 178. Para un punto de vista más crítico de $R^2$, véase S. Cameron, "Why is the *R* Squared Adjusted Reported?", *Journal of Quantitative Economics*, vol. 9, núm. 1, enero de 1993, pp. 183-186. Argumenta que "[$R^2$] NO es un estadístico de prueba y parece que no hay una justificación intuitiva clara para usarlo como estadístico descriptivo. Finalmente, debe quedar claro que no es una herramienta eficaz para la prevención de la búsqueda exhaustiva de datos" (p. 186).

A pesar de esta sugerencia, es la $R^2$ ajustada, como aparece en (7.8.4), la que se utiliza en la mayoría de los paquetes estadísticos junto con $R^2$ convencional. Se aconseja al lector tratar $\bar{R}^2$ como cualquier otro estadístico de resumen.

A propósito, para la regresión sobre la mortalidad infantil (7.6.2), el lector debe verificar que $\bar{R}^2$ es 0.6981, pero sin olvidar que en este ejemplo $(n - 1) = 63$ y que $(n - k) = 60$. Como era de esperarse, $\bar{R}^2$ de 0.6981 es menor que $R^2$ de 0.7077.

Además de $R^2$ y $R^2$ ajustada como medidas de bondad de ajuste, a menudo se utilizan otros criterios para juzgar la bondad de un modelo de regresión. Dos de estos son el **criterio de información de Akaike** y el **criterio de predicción de Amemiya**, con los cuales se escogen modelos que compiten. Analizaremos estos criterios cuando veamos el problema de selección de modelos en mayor detalle en un capítulo posterior (el 13).

## Comparación de dos valores de $R^2$

Es de crucial importancia señalar que, al comparar dos modelos con base en el coeficiente de determinación, ajustado o no, *el tamaño de la muestra n y la variable dependiente deben ser los mismos*; las variables explicativas pueden adoptar cualquier forma. Así, para los modelos

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(7.8.6)}$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \qquad \textbf{(7.8.7)}$$

no pueden compararse los términos $R^2$ calculados. La razón es la siguiente: por definición, $R^2$ mide la proporción de la variación en la variable dependiente explicada por la(s) variable(s) explicativa(s). Por consiguiente, en (7.8.6), el $R^2$ mide la proporción de la *variación en* ln $Y$ explicada por $X_2$ y $X_3$, mientras que en (7.8.7), mide la proporción de la *variación en* $Y$, y las dos no son la misma variable: Como vimos en el capítulo 6, un cambio en ln $Y$ da un cambio relativo o proporcional en $Y$, mientras que un cambio en $Y$ da un cambio absoluto. Por consiguiente, $\text{var}\,\hat{Y}_i/\text{var}\,Y_i$ no es igual a $\text{var}\,(\widehat{\ln Y_i})/\text{var}\,(\ln Y_i)$; es decir, los dos coeficientes de determinación no son lo mismo.[13]

Entonces, ¿cómo comparar las $R^2$ de los dos modelos cuando la regresada no está en la misma forma? Para responder, consideremos primero un ejemplo numérico.

---

[13] De la definición de $R^2$, sabemos que

$$1 - R^2 = \frac{\text{SCR}}{\text{TSS}} = \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

para el modelo lineal y

$$1 - R^2 = \frac{\sum \hat{u}_i^2}{\sum (\ln Y_i - \overline{\ln Y})^2}$$

para el modelo log. Como los denominadores en el lado derecho de estas expresiones son diferentes, no podemos comparar los dos términos $R^2$ directamente.

Como se muestra en el ejemplo 7.2, para la especificación lineal, SCR = 0.1491 (la suma de cuadrados de residuos del consumo de café), y para la especificación log-lineal, SCR = 0.0226 (la suma de cuadrados de residuos del logaritmo del consumo de café). Estos residuos son de diferentes órdenes de magnitud y por tanto no son directamente comparables.

**EJEMPLO 7.2**
*Consumo de café en Estados Unidos, 1970-1980*

Considere los datos de la tabla 7.1, los cuales se refieren al consumo de tazas de café por día ($Y$) y el precio al menudeo del café ($X$) en Estados Unidos de 1970 a 1980. Al aplicar MCO a los datos se obtienen los siguientes resultados de la regresión:

$$\begin{aligned}\hat{Y}_t &= 2.6911 \quad - \; 0.4795X_t \\ \text{ee} &= (0.1216) \quad (0.1140) \qquad \text{SCR} = 0.1491;\ r^2 = 0.6628\end{aligned} \tag{7.8.8}$$

Los resultados tienen sentido en el contexto económico: conforme se incrementa el precio del café, en promedio, su consumo disminuye casi media taza al día. El valor $r^2$ de más o menos 0.66 indica que el precio del café explica casi 66% de la variación en el consumo del café. El lector puede verificar con facilidad que la pendiente del coeficiente es estadísticamente significativa.

A partir de los mismos datos, el siguiente modelo de doble logaritmo, o elasticidad constante, se estima:

$$\begin{aligned}\widehat{\ln Y_t} &= 0.7774 \quad - \; 0.2530 \ln X_t \\ \text{ee} &= (0.0152) \quad (0.0494) \qquad \text{SCR} = 0.0226;\ r^2 = 0.7448\end{aligned} \tag{7.8.9}$$

Como es un modelo de doble logaritmo, el coeficiente de la pendiente proporciona un estimado directo del coeficiente de elasticidad del precio. En el ejemplo presente, indica que si el precio del café por libra se incrementa 1%, en promedio, su consumo diario disminuye casi 0.25%. Recuerde que, en el modelo lineal (7.8.8), el coeficiente de la pendiente sólo señala la tasa de cambio del consumo del café respecto del precio. (¿Cómo estimará la elasticidad-precio en el modelo lineal?) El valor $r^2$ de casi 0.74 significa que 74% de la variación en el logaritmo de la demanda de café se explica por la variación en el logaritmo del precio del café.

Como el valor $r^2$ (0.6628) del modelo lineal es menor que el valor $r^2$ de 0.7448 del modelo lineal logarítmico, se presenta la tentación de elegir este último modelo debido al alto valor de $r^2$. Sin embargo, por las razones expresadas, no es posible hacerlo así. No obstante, si desea comparar ambos valores $r^2$, puede proceder de la siguiente manera:

**TABLA 7.1**
**Consumo de café en Estados Unidos ($Y$) respecto del precio promedio real al menudeo ($X$),* 1970-1980**

Fuente: Los datos de $Y$ provienen del *Summary of National Coffee Drinking Study*, Data Group, Elkins Park, Pensilvania, 1981, y los datos sobre $X$ nominal (es decir, $X$ en precios corrientes), de *Nielsen Food Index*, A. C. Nielsen, Nueva York, 1981.

El autor está en deuda con Scott E. Sandberg por la recopilación de los datos.

| Año | $Y$, Tazas diarias por persona | $X$, $ por libra |
|---|---|---|
| 1970 | 2.57 | 0.77 |
| 1971 | 2.50 | 0.74 |
| 1972 | 2.35 | 0.72 |
| 1973 | 2.30 | 0.73 |
| 1974 | 2.25 | 0.76 |
| 1975 | 2.20 | 0.75 |
| 1976 | 2.11 | 1.08 |
| 1977 | 1.94 | 1.81 |
| 1978 | 1.97 | 1.39 |
| 1979 | 2.06 | 1.20 |
| 1980 | 2.02 | 1.17 |

**Nota:* El precio nominal se dividió entre el IPC para alimentos y bebidas, 1967 = 100.

1. Obtenga $\widehat{\ln Y_t}$ de (7.8.9) para cada observación; es decir, encuentre el valor estimado de cada observación a partir de este modelo. Tome el antilogaritmo de esos valores y después calcule $r^2$ entre dichos valores del antilogaritmo y la verdadera $Y_t$ de la manera señalada por la ecuación (3.5.14). Este valor $r^2$ es comparable con el valor $r^2$ del modelo lineal (7.8.8).
2. *Otra forma* es suponer que todos los valores $Y$ son positivos, en cuyo caso calcule los logaritmos de los valores $Y$, ln $Y$. Obtenga los valores estimados $Y$, $\hat{Y}_t$, del modelo lineal (7.8.8), calcule los logaritmos de dichos valores estimados $Y$ (es decir, ln $\hat{Y}_t$) y calcule la $r^2$ entre (ln $Y_t$) y (ln $\hat{Y}_t$) como en la ecuación (3.5.14). Este valor $r^2$ es comparable con el valor $r^2$ obtenido mediante (7.8.9).

Para el ejemplo del café, en la tabla 7.2 se presentan los datos originales necesarios para calcular las $r^2$ comparables. A fin de comparar el valor $r^2$ del modelo lineal (7.8.8) con el de (7.8.9), primero obtenemos el logaritmo de ($\hat{Y}_t$) [dado en la columna (6) de la tabla 7.2], luego calculamos el logaritmo de los valores reales $Y$ [dados en la columna (5) de la tabla] y por último calculamos $r^2$ entre estos dos conjuntos de valores mediante la ecuación (3.5.14). El resultado es un valor $r^2$ de 0.6779, el cual ahora se puede comparar con el valor $r^2$ de 0.7448 del modelo log-lineal. La diferencia entre ambos valores $r^2$ es aproximadamente 0.07.

Por otra parte, si deseamos comparar el valor $r^2$ del modelo log-lineal con el obtenido del modelo lineal, estimamos $\widehat{\ln Y_t}$ para cada observación de (7.8.9) [dadas en la columna (3) de la tabla], obtenemos sus valores antilog [dados en la columna (4) de la tabla] y por último calculamos $r^2$ entre estos valores antilog y los valores reales de $Y$ observados mediante la fórmula (3.5.14). Esto da a $r^2$ un valor de 0.7187, el cual es un poco superior al valor obtenido del modelo lineal (7.8.8) de 0.6628.

Con cualquier método, parece que el modelo log-lineal ofrece un ajuste ligeramente mejor.

**TABLA 7.2**
**Datos básicos para comparar dos valores de $R^2$**

| Año | $Y_t$ (1) | $\hat{Y}_t$ (2) | $\widehat{\ln Y_t}$ (3) | Antilog de $\widehat{\ln Y_t}$ (4) | ln $Y_t$ (5) | ln ($\hat{Y}_t$) (6) |
|---|---|---|---|---|---|---|
| 1970 | 2.57 | 2.321887 | 0.843555 | 2.324616 | 0.943906 | 0.842380 |
| 1971 | 2.50 | 2.336272 | 0.853611 | 2.348111 | 0.916291 | 0.848557 |
| 1972 | 2.35 | 2.345863 | 0.860544 | 2.364447 | 0.854415 | 0.852653 |
| 1973 | 2.30 | 2.341068 | 0.857054 | 2.356209 | 0.832909 | 0.850607 |
| 1974 | 2.25 | 2.326682 | 0.846863 | 2.332318 | 0.810930 | 0.844443 |
| 1975 | 2.20 | 2.331477 | 0.850214 | 2.340149 | 0.788457 | 0.846502 |
| 1976 | 2.11 | 2.173233 | 0.757943 | 2.133882 | 0.746688 | 0.776216 |
| 1977 | 1.94 | 1.823176 | 0.627279 | 1.872508 | 0.662688 | 0.600580 |
| 1978 | 1.97 | 2.024579 | 0.694089 | 2.001884 | 0.678034 | 0.705362 |
| 1979 | 2.06 | 2.115689 | 0.731282 | 2.077742 | 0.722706 | 0.749381 |
| 1980 | 2.02 | 2.130075 | 0.737688 | 2.091096 | 0.703098 | 0.756157 |

*Notas:* Columna (1): Valores reales de $Y$ de la tabla 7.1.
Columna (2): Valores estimados de $Y$ del modelo lineal (7.8.8).
Columna (3): Valores estimados de log $Y$ del modelo doble-log (7.8.9).
Columna (4): Antilog de valores de la columna (3).
Columna (5): Valores log de $Y$ en la columna (1).
Columna (6): Valores log de $\hat{Y}_t$ en la columna (2).

## Asignación de $R^2$ entre regresoras

Retomemos el ejemplo de la mortalidad infantil. Vimos, con base en (7.6.2), que las dos regresoras, PIBPC y TAM, explican 0.7077 o 70.77% de la variación de la mortalidad infantil. Pero ahora considere la regresión (7.7.2), donde se eliminó la variable TAM y como consecuencia el valor $r^2$ disminuyó hasta 0.1662. ¿Lo anterior significa que la diferencia en el valor $r^2$ de 0.5415 (0.7077 − 0.1662) se atribuye a la variable omitida, TAM? Por otro lado, si considera la regresión (7.7.3) en la que se quitó la variable PIB, el valor $r^2$ disminuye hasta 0.6696. ¿Significa que la diferencia en el valor $r^2$ de 0.0381 (0.7077 − 0.6696) se debe a la variable omitida, PIBPC?

Por tanto, la pregunta es: ¿podemos asignar la $R^2$ múltiple de 0.7077 entre las dos regresoras, PIBPC y TAM, de esta forma? Por desgracia, no, pues la asignación depende del orden de introducción de las regresoras, como acabamos de ilustrar. Parte de este problema radica en que las dos regresoras están correlacionadas, pues el coeficiente de correlación entre ambas es igual a 0.2685 (verifique a partir de los datos proporcionados en la tabla 6.4). En la mayor parte del trabajo con varias regresoras, la correlación entre ambas constituye un problema común. Por supuesto, dicho problema sería mucho más grave si existiese una perfecta colinealidad entre las regresoras.

El mejor consejo práctico es que no tiene mucho sentido tratar de asignar el valor $R^2$ a sus regresoras constituyentes.

## El "juego" de maximizar $\bar{R}^2$

Para concluir esta sección vale la pena una advertencia: algunas veces, los investigadores intentan maximizar $\bar{R}^2$, es decir, escogen el modelo que da la $\bar{R}^2$ más elevada. Pero esto puede ser peligroso, pues, en el análisis de regresión, el objetivo no es obtener una $\bar{R}^2$ elevada *per se,* sino más bien obtener estimados confiables de los verdaderos coeficientes de regresión poblacional que permitan realizar inferencia estadística sobre ellos. En el análisis empírico no es inusual obtener una $\bar{R}^2$ muy elevada, sino descubrir que algunos de los coeficientes de regresión no son estadísticamente significativos o muestran signos contrarios a los esperados *a priori*. Por consiguiente, el investigador debe preocuparse más por la pertinencia lógica o teórica de las variables explicativas para la variable dependiente y por su significancia estadística. Si en este proceso obtenemos una $\bar{R}^2$ elevada, muy bien; por otra parte, si $\bar{R}^2$ es baja, esto no significa que el modelo sea necesariamente malo.[14]

De hecho Goldberger, quien es muy crítico sobre el papel de $R^2$, afirmó:

> Desde nuestra perspectiva, $R^2$ tiene un papel muy modesto en el análisis de regresión, y es una medida de la bondad del ajuste de una regresión lineal por MC [mínimos cuadrados] muestral en un cuerpo de datos. Nada en el modelo de RC [MCRL] exige que $R^2$ sea elevada. Por tanto, una $R^2$ elevada no es evidencia en favor del modelo y una $R^2$ baja no es evidencia en su contra.
>
> En realidad, lo más importante sobre $R^2$ es que no es importante en el modelo de RC. El modelo de RC tiene que ver con parámetros en una población, no con la bondad de ajuste en la muestra…

[14] Algunos autores desearían reducir el acento en el uso de $R^2$ como medida de bondad del ajuste, al igual que para comparar dos o más valores de $R^2$. Véase Christopher H. Achen, *Interpreting and Using Regression*, Sage Publications, Beverly Hills, California, 1982, pp. 58-67, y C. Granger y P. Newbold, "$R^2$ and the Transformation of Regression Variables", *Journal of Econometrics*, vol. 4, 1976, pp. 205-210. A propósito, la práctica de seleccionar un modelo con base en la $R^2$ más elevada, una especie de búsqueda exhaustiva, introduce lo que se conoce como **sesgo de preprueba**, que puede destruir algunas propiedades de los estimadores de MCO del modelo clásico de regresión lineal. Sobre este tema, el lector puede consultar George G. Judge, Carter R. Hill, William E. Griffiths, Helmut Lütkepohl y Tsoun-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley, Nueva York, 1982, capítulo 21.

> Si se insiste en una medida del éxito predictivo (o más bien el fracaso), bastaría $\sigma^2$: después de todo, el parámetro $\sigma^2$ es el error de predicción esperado al cuadrado que resultaría si se empleara la FEC [FRP] poblacional como predictora. Esto es, el error estándar de predicción elevado al cuadrado para valores relevantes de $x$ [regresoras] puede ser informativo.[15]

## 7.9 La función de producción Cobb-Douglas: más sobre la forma funcional

En la sección 6.4 demostramos cómo convertir, mediante transformaciones apropiadas, las relaciones no lineales en relaciones lineales, de forma que se facilite trabajar dentro del marco del MCRL. Las diversas transformaciones analizadas allí en el contexto del caso de dos variables se amplían sin dificultad a los modelos de regresión múltiple. En esta sección demostramos estas transformaciones con una extensión multivariada del modelo log-lineal de dos variables; hay más en los ejercicios y en los ejemplos ilustrativos en el resto de este libro. El ejemplo específico es la conocida **función de producción Cobb-Douglas** de la teoría de producción.

La función de producción Cobb-Douglas, en su forma estocástica, se expresa como

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{7.9.1}$$

donde $Y =$ producción
$X_2 =$ insumo trabajo
$X_3 =$ insumo capital
$u =$ término de perturbación estocástica
$e =$ base del logaritmo natural

De la ecuación (7.9.1) es claro que la relación entre la producción y los dos insumos es no lineal. Sin embargo, si transformamos este modelo, mediante la función logaritmo, tenemos:

$$\begin{aligned} \ln Y_i &= \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \end{aligned} \tag{7.9.2}$$

donde $\beta_0 = \ln \beta_1$.

Escrito de esta forma, el modelo es lineal en los parámetros $\beta_0$, $\beta_2$ y $\beta_3$, y por consiguiente es un modelo de regresión lineal. Observe, sin embargo, que es no lineal en las variables $Y$ y $X$, aunque sí lo es en sus logaritmos. En resumen, (7.9.2) es un m*odelo log-log, doble-log* o *log-lineal*, el equivalente en la regresión múltiple al modelo log-lineal con dos variables (6.5.3).

Las propiedades de la función de producción Cobb-Douglas son bien conocidas:

1. $\beta_2$ es la elasticidad (parcial) de la producción respecto del insumo trabajo, es decir, mide el cambio porcentual en la producción debido a una variación de 1% en el insumo trabajo, con el insumo capital constante (véase el ejercicio 7.9).

2. De igual forma, $\beta_3$ es la elasticidad (parcial) de la producción respecto del insumo capital, con el insumo trabajo constante.

3. La suma $(\beta_2 + \beta_3)$ da información sobre los *rendimientos a escala*, es decir, la respuesta de la producción a un cambio proporcional en los insumos. Si esta suma es 1, existen *rendimientos constantes a escala*, es decir, la duplicación de los insumos duplica la producción, la triplicación

[15] Arthur S. Goldberger, *op. cit.*, pp. 177-178.

de los insumos la triplica, y así sucesivamente. Si la suma es menor que 1, existen *rendimientos decrecientes a escala*: al duplicar los insumos, la producción crece en menos del doble. Por último, si la suma es mayor que 1, hay *rendimientos crecientes a escala*; la duplicación de los insumos aumenta la producción en más del doble.

Antes de continuar, observe que siempre que se tenga un modelo de regresión log-lineal con cualquier número de variables, el coeficiente de cada variable $X$ mide la elasticidad (parcial) de la variable dependiente $Y$ respecto de esa variable. Así, si se tiene un modelo log-lineal con $k$ variables:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \cdots + \beta_k \ln X_{ki} + u_i \quad (7.9.3)$$

cada uno de los coeficientes de regresión (parcial), $\beta_2$ hasta $\beta_k$, es la elasticidad (parcial) de $Y$ respecto de las variables $X_2$ hasta $X_k$.[16]

**EJEMPLO 7.3**
*Valor agregado, horas de mano de obra y aportación de capital en el sector manufacturero*

Para ilustrar la función de producción Cobb-Douglas se obtuvieron los datos de la tabla 7.3, referentes al sector manufacturero de los 50 estados de Estados Unidos y Washington, D.C., para 2005.

Si el modelo (7.9.2) satisface los supuestos del modelo clásico de regresión lineal,[17] obtenemos la siguiente regresión por el método de MCO (véase el listado de computadora en el apéndice 7A, sección 7A.5):

**TABLA 7.3**
**Valor agregado, horas de mano de obra y aportación de capital en el sector manufacturero de Estados Unidos, 2005**

Fuente: *2005 Annual Survey of Manufacturers*, Sector 31: Estadísticas complementarias de Estados Unidos.

| Área | Producción Valor agregado (miles de dólares) *Y* | Insumo mano de obra Horas de trabajo (miles) *X2* | Insumo capital Inversión de capital (miles de dólares) *X3* |
|---|---|---|---|
| Alabama | 38 372 840 | 424 471 | 2 689 076 |
| Alaska | 1 805 427 | 19 895 | 57 997 |
| Arizona | 23 736 129 | 206 893 | 2 308 272 |
| Arkansas | 26 981 983 | 304 055 | 1 376 235 |
| California | 217 546 032 | 1 809 756 | 13 554 116 |
| Colorado | 19 462 751 | 180 366 | 1 790 751 |
| Connecticut | 28 972 772 | 224 267 | 1 210 229 |
| Delaware | 14 313 157 | 54 455 | 421 064 |
| Distrito de Columbia | 159 921 | 2 029 | 7 188 |
| Florida | 47 289 846 | 471 211 | 2 761 281 |
| Georgia | 63 015 125 | 659 379 | 3 540 475 |
| Hawaii | 1 809 052 | 17 528 | 146 371 |
| Idaho | 10 511 786 | 75 414 | 848 220 |
| Illinois | 105 324 866 | 963 156 | 5 870 409 |
| Indiana | 90 120 459 | 835 083 | 5 832 503 |
| Iowa | 39 079 550 | 336 159 | 1 795 976 |
| Kansas | 22 826 760 | 246 144 | 1 595 118 |
| Kentucky | 38 686 340 | 384 484 | 2 503 693 |
| Louisiana | 69 910 555 | 216 149 | 4 726 625 |

[16] Para ver esto, diferencie parcialmente la ecuación (7.9.3) respecto del log de cada variable $X$. Por consiguiente, $\partial \ln Y/\partial \ln X_2 = (\partial Y/\partial X_2)(X_2/Y) = \beta_2$, que, por definición, es la elasticidad de $Y$ respecto de $X_2$ y $\partial \ln Y/\partial \ln X_3 = (\partial Y/\partial X_3)(X_3/Y) = \beta_3$, que es la elasticidad de $Y$ respecto de $X_3$, y así sucesivamente.

[17] Observe que en la función de producción Cobb-Douglas (7.9.1) se presentó el término de error estocástico en una forma especial que, en la transformación logarítmica resultante, ingrese en la forma lineal usual. Sobre este tema, véase la sección 6.9.

| | | | |
|---|---|---|---|
| Maine | 7 856 947 | 82 021 | 415 131 |
| Maryland | 21 352 966 | 174 855 | 1 729 116 |
| Massachusetts | 46 044 292 | 355 701 | 2 706 065 |
| Michigan | 92 335 528 | 943 298 | 5 294 356 |
| Minnesota | 48 304 274 | 456 553 | 2 833 525 |
| Mississippi | 17 207 903 | 267 806 | 1 212 281 |
| Missouri | 47 340 157 | 439 427 | 2 404 122 |
| Montana | 2 644 567 | 24 167 | 334 008 |
| Nebraska | 14 650 080 | 163 637 | 627 806 |
| Nevada | 7 290 360 | 59 737 | 522 335 |
| New Hampshire | 9 188 322 | 96 106 | 507 488 |
| New Jersey | 51 298 516 | 407 076 | 3 295 056 |
| New Mexico | 20 401 410 | 43 079 | 404 749 |
| New York | 87 756 129 | 727 177 | 4 260 353 |
| North Carolina | 101 268 432 | 820 013 | 4 086 558 |
| North Dakota | 3 556 025 | 34 723 | 184 700 |
| Ohio | 124 986 166 | 1 174 540 | 6 301 421 |
| Oklahoma | 20 451 196 | 201 284 | 1 327 353 |
| Oregon | 34 808 109 | 257 820 | 1 456 683 |
| Pennsylvania | 104 858 322 | 944 998 | 5 896 392 |
| Rhode Island | 6 541 356 | 68 987 | 297 618 |
| South Carolina | 37 668 126 | 400 317 | 2 500 071 |
| South Dakota | 4 988 905 | 56 524 | 311 251 |
| Tennessee | 62 828 100 | 582 241 | 4 126 465 |
| Texas | 172 960 157 | 1 120 382 | 11 588 283 |
| Utah | 15 702 637 | 150 030 | 762 671 |
| Vermont | 5 418 786 | 48 134 | 276 293 |
| Virginia | 49 166 991 | 425 346 | 2 731 669 |
| Washington | 46 164 427 | 313 279 | 1 945 860 |
| West Virginia | 9 185 967 | 89 639 | 685 587 |
| Wisconsin | 66 964 978 | 694 628 | 3 902 823 |
| Wyoming | 2 979 475 | 15 221 | 361 536 |

$$\begin{aligned}\widehat{\ln Y_i} &= 3.8876 + 0.4683 \ln X_{2i} + 0.5213 \ln X_{3i} \\ &\quad (0.3962) \quad (0.0989) \qquad (0.0969) \\ t &= (9.8115) \quad (4.7342) \qquad (5.3803) \\ &\qquad R^2 = 0.9642 \qquad \text{gl} = 48 \\ &\qquad \bar{R}^2 = 0.9627\end{aligned} \tag{7.9.4}$$

De la ecuación (7.9.4), vemos que, en el sector manufacturero de Estados Unidos durante 2005, las elasticidades de la producción respecto del trabajo y el capital fueron 0.4683 y 0.5213, respectivamente. En otras palabras, en los 50 estados de Estados Unidos y el Distrito de Columbia, manteniendo constante el insumo capital, un incremento de 1% en el insumo trabajo provocó, en promedio, un incremento de cerca de 0.47% en la producción. En forma similar, manteniendo constante el insumo trabajo, un incremento de 1% en el insumo capital generó, en promedio, un incremento de cerca de 0.52% en la producción. Sumamos las dos elasticidades de la producción y obtenemos 0.99, que da el valor del parámetro de rendimientos a escala. Como es evidente, el sector manufacturero de los 50 estados de Estados Unidos y el Distrito de Columbia se caracterizó por rendimientos constantes a escala.[18]

Desde el punto de vista puramente estadístico, la línea de regresión estimada se ajusta muy bien a los datos. El valor $R^2$ de 0.9642 significa que cerca de 96% de la variación en el (log de la) producción se explica por el (log del) trabajo y el (log del) capital. En el capítulo 8 veremos cómo utilizar los errores estándar estimados para probar hipótesis sobre los "verdaderos" valores de los parámetros de la función de producción Cobb-Douglas en el sector manufacturero de la economía estadounidense.

## 7.10 Modelos de regresión polinomial

Ahora consideraremos una clase de modelos de regresión múltiple, los **modelos de regresión polinomial,** de amplio uso en la investigación econométrica relacionada con funciones de costo y de producción. Al introducir estos modelos, ampliamos la gama de modelos a todos los que se aplica fácilmente el modelo clásico de regresión lineal.

Para ordenar las ideas, considere la figura 7.1 que relaciona el costo marginal (CM) de corto plazo de la producción de un bien ($Y$) con el nivel de su producción ($X$). La curva de CM de la figura, la curva con forma de U de los libros de texto, muestra que la relación entre CM y producción es no lineal. Si se cuantificara esta relación a partir de los puntos dispersos dados, ¿cómo se haría? En otras palabras, ¿qué tipo de modelo econométrico expresa la naturaleza primero decreciente y luego creciente del costo marginal?

Geométricamente, la curva CM de la figura 7.1 representa una *parábola*. Matemáticamente, la parábola está representada por la siguiente ecuación:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{7.10.1}$$

que se denomina una *función cuadrática* o, más generalmente, un *polinomio de segundo grado* en la variable $X$; la mayor potencia de $X$ representa el grado del polinomio (si se agregara $X^3$ a la función anterior, sería un polinomio de tercer grado, y así sucesivamente).

La versión estocástica de (7.10.1) se escribe así:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \tag{7.10.2}$$

que se denomina *regresión polinomial de segundo grado*.

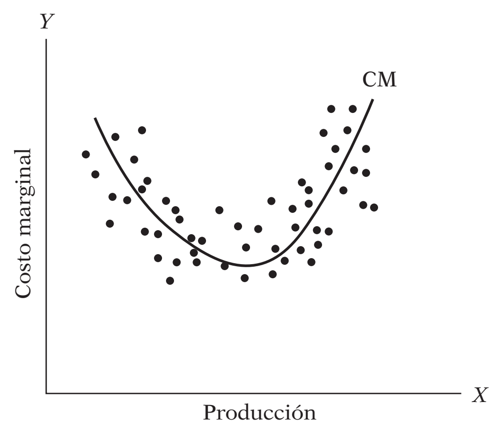
La *regresión polinomial de grado k* general puede escribirse así:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i \tag{7.10.3}$$

Tenga en cuenta que, en estos tipos de regresiones polinomiales, sólo hay una variable explicativa al lado derecho, pero aparece elevada a distintas potencias, convirtiéndolas en modelos de regresión múltiple. A propósito, observe que si se supuso que $X_i$ es fija o no estocástica, los términos de $X_i$ elevados a alguna potencia también se hacen fijos o no estocásticos.

¿Presentan estos modelos problemas especiales de estimación? Como el polinomio de segundo grado (7.10.2) o el polinomio de grado $k$ (7.10.13) son lineales en los parámetros, las $\beta$ se estiman mediante las metodologías usuales de MCO o MV. Pero, ¿qué sucede con el problema

**FIGURA 7.1**
Curva de costo marginal en forma de U.



de colinealidad? ¿Acaso las diferentes $X$ no están altamente correlacionadas puesto que todas son potencias de $X$? Sí, pero recuerde que todos los términos como $X^2$, $X^3$, $X^4$, etc., son funciones no lineales de $X$ y, por consiguiente, en términos estrictos, no violan el supuesto de no multicolinealidad. En resumen, es posible estimar modelos de regresión polinomial mediante las técnicas estudiadas en este capítulo sin que se presenten nuevos problemas de estimación.

**EJEMPLO 7.4**
*Estimación de la función de costo total*

Como ejemplo de regresión polinomial, considere los datos de la tabla 7.4 sobre producción de un bien y su costo de producción total en el corto plazo. ¿Qué tipo de modelo de regresión ajusta estos datos? Para este fin, trace primero el diagrama de dispersión, que se muestra en la figura 7.2.

De esta figura es claro que la relación entre el costo total y la producción semeja una curva en forma de S alargada; observe cómo la curva de costo total primero aumenta poco a poco y luego lo hace rápido, como lo establece la conocida ley de *rendimientos decrecientes*. Esta forma de S de la curva de costo total se representa por el siguiente *polinomio cúbico* o *de tercer grado*:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad \textbf{(7.10.4)}$$

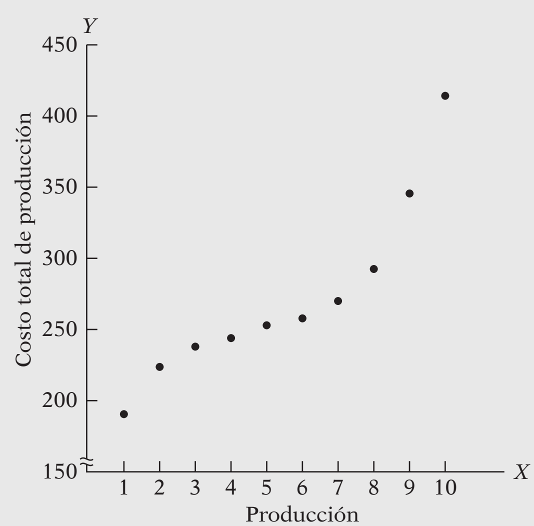
donde $Y$ = costo total y $X$ = producción.

En virtud de los datos de la tabla 7.4, aplicamos el método de MCO para estimar los parámetros de (7.10.4). Pero, antes de hacerlo, vea lo que la teoría económica indica sobre la función cúbica de costo de corto plazo (7.10.4). La teoría elemental de precios muestra que, en el corto plazo, las curvas de costo marginal de producción (CM) y de costo promedio (CP) en general tienen forma de U (al principio), a medida que la producción aumenta tanto el CM como el CP decrecen, pero, después de un nivel dado de producción, ambas vuelven a aumentar, de nuevo como consecuencia de la ley de rendimientos decrecientes. Esto se aprecia en la figura 7.3 (véase también la figura 7.1). Y, como las curvas de CM y de CP se derivan de la curva de costo total, la naturaleza de estas curvas en forma de U impone algunas restricciones sobre los parámetros de la curva de costo total (7.10.4). De hecho, puede mostrarse que los parámetros

**TABLA 7.4**
**Costo total ($Y$) y producción ($X$)**

| Producción | Costo total, $ |
|---|---|
| 1 | 193 |
| 2 | 226 |
| 3 | 240 |
| 4 | 244 |
| 5 | 257 |
| 6 | 260 |
| 7 | 274 |
| 8 | 297 |
| 9 | 350 |
| 10 | 420 |

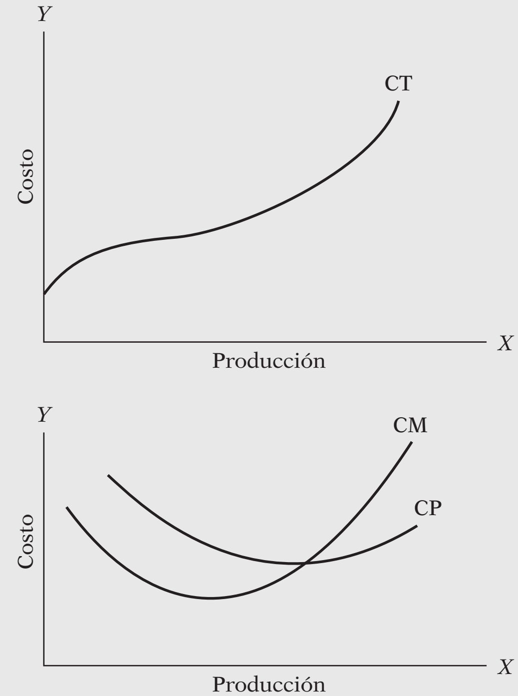
**FIGURA 7.2** Curva de costo total.




(*continúa*)

**EJEMPLO 7.4**
(*continuación*)

**FIGURA 7.3** Funciones de costo de corto plazo.




de (7.10.4) deben satisfacer las siguientes restricciones si se desea observar las curvas de costo marginal y promedio de corto plazo en la forma típica de U:[18]

1. $\beta_0$, $\beta_1$ y $\beta_3 > 0$
2. $\beta_2 < 0$ **(7.10.5)**
3. $\beta_2^2 < 3\beta_1\beta_3$

Toda esta exposición teórica puede parecer un poco tediosa. Pero este conocimiento es en extremo útil cuando se examinan los resultados empíricos, pues, si éstos no concuerdan con las expectativas *a priori*, entonces, suponiendo que no se cometió un error de especificación (es decir, que se escogió el modelo erróneo), se tendrá que modificar la teoría, o buscar una nueva y reiniciar la investigación empírica desde el principio. Pero, como comentamos en la introducción, ésta es la naturaleza de toda investigación empírica.

**Resultados empíricos.** Cuando la regresión polinomial de tercer grado se ajustó a los datos de la tabla 7.4, obtuvimos los siguientes resultados:

$$\hat{Y}_i = 141.7667 + 63.4776X_i - 12.9615X_i^2 + 0.9396X_i^3$$
$$(6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591) \quad R^2 + 0.9983 \qquad \textbf{(7.10.6)}$$

(*Nota:* Las cifras entre paréntesis son los errores estándar estimados.) Aunque examinaremos la significancia estadística de estos resultados en el siguiente capítulo, el lector puede verificar que corresponden a las expectativas teóricas de (7.10.5). Como ejercicio para el lector queda la tarea de interpretar la regresión (7.10.6).

[18] Véase Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, 3a. ed., McGraw-Hill, Nueva York, 1984, pp. 250-252.

**EJEMPLO 7.5**
*Tasa de crecimiento del PIB y PIB per cápita relativo para 2007 en 190 países (en miles de millones de dólares de 2000)*

Fuente: Indicadores de desarrollo mundial del Banco Mundial, ajustados a la base de 2000, y valores estimados y proyectados obtenidos por el Economic Research Service

Como otro ejemplo económico del modelo de regresión polinomial, considere los siguientes resultados de regresión:

$$\begin{aligned}\widehat{\text{TPIB}}_i &= 5.5347 - 5.5788\ \text{PIBR} + 2.8378\ \text{PIBR}^2 \\ \text{ee} &= (0.2435) \quad (1.5995) \qquad\quad (1.4391) \\ R^2 &= 0.1092 \qquad R^2 \text{ ajs} = 0.0996\end{aligned} \tag{7.10.7}$$

donde TPIB = tasa de crecimiento del PIB, porcentaje en 2007, y PIBR = PIB *per cápita* relativo en 2007 (porcentaje del PIB *per cápita* de Estados Unidos en 2007). La $R^2$ ajustada ($R^2$ ajs) indica que, después de considerar el número de regresoras, el modelo sólo explica alrededor de 9.96% de la variación en el TPIB. Incluso la $R^2$ sin ajustar de 0.1092 parece baja. Lo anterior puede parecer un valor muy desalentador, pero, como veremos en el siguiente capítulo, tales valores bajos de $R^2$ suelen encontrarse en datos transversales con un gran número de observaciones. Además, incluso una $R^2$ en apariencia baja puede ser estadísticamente significativa (es decir, distinta de cero), como mostraremos en el próximo capítulo.

## *7.11 Coeficientes de correlación parcial

### Explicación de los coeficientes de correlación simple y parcial

En el capítulo 3 presentamos el coeficiente de correlación $r$ como medida del grado de asociación lineal entre dos variables. Para el modelo de regresión con tres variables podemos calcular tres coeficientes de correlación: $r_{12}$ (correlación entre $Y$ y $X_2$), $r_{13}$ (coeficiente de correlación entre $Y$ y $X_3$) y $r_{23}$ (coeficiente de correlación entre $X_2$ y $X_3$); observe que el subíndice 1 representa a $Y$ por conveniencia notacional. Estos coeficientes de correlación se denominan **coeficientes de correlación bruta** o **simple**, o **coeficientes de correlación de orden cero**, y se calculan a partir de la definición del coeficiente de correlación dada en (3.5.13).

Pero consideremos ahora esta interrogante: ¿podemos decir en realidad que $r_{12}$ mide el "verdadero" grado de asociación (lineal) entre $Y$ y $X_2$ cuando existe una tercera variable $X_3$ que puede estar asociada a ellas? Esta pregunta es análoga a la siguiente: suponga que el verdadero modelo de regresión es (7.1.1) pero omitimos del modelo la variable $X_3$, y sólo hacemos la regresión $Y$ sobre $X_2$ para obtener el coeficiente de la pendiente de, por ejemplo, $b_{12}$. ¿Será igual este coeficiente al verdadero coeficiente $\beta_2$ si, para empezar, se estimara el modelo (7.1.1)? La respuesta debe ser clara a partir del análisis en la sección 7.7. En general, $r_{12}$ tal vez no refleje el verdadero grado de asociación entre $Y$ y $X_2$ en presencia de $X_3$. De hecho, es probable que dé una falsa impresión de la naturaleza de la asociación entre $Y$ y $X_2$, como demostraremos en breve. Por consiguiente, lo que se necesita es un coeficiente de correlación que sea independiente de la influencia, si hay alguna, de $X_3$ sobre $X_2$ y $Y$. Dicho coeficiente de correlación se obtiene y se conoce apropiadamente como **coeficiente de correlación parcial**. En cuanto concepto, es similar al coeficiente de regresión parcial. Definimos

$r_{12.3}$ = coeficiente de correlación parcial entre $Y$ y $X_2$, manteniendo $X_3$ constante

$r_{13.2}$ = coeficiente de correlación parcial entre $Y$ y $X_3$, manteniendo $X_2$ constante

$r_{23.1}$ = coeficiente de correlación parcial entre $X_2$ y $X_3$, manteniendo $Y$ constante

* Opcional.

Estas correlaciones parciales se obtienen con facilidad a partir de los coeficientes de correlación simples o de orden cero, de la siguiente forma (para las pruebas, véase los ejercicios):[19]

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} \tag{7.11.1}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{23}^2\right)}} \tag{7.11.2}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{13}^2\right)}} \tag{7.11.3}$$

Las correlaciones parciales de las ecuaciones (7.11.1) a (7.11.3) se denominan **coeficientes de correlación de primer orden**. Por *orden* se quiere decir el número de subíndices secundarios. Así $r_{12.34}$ sería el coeficiente de correlación de orden dos, $r_{12.345}$ sería el coeficiente de correlación de orden tres, y así sucesivamente. Como ya vimos, $r_{12}$, $r_{13}$ y las siguientes se denominan *correlaciones simples* o *de orden cero*. La interpretación de $r_{12.34}$, por ejemplo, es que éste da el coeficiente de correlación entre $Y$ y $X_2$, manteniendo constantes $X_3$ y $X_4$.

## Interpretación de los coeficientes de correlación simple y parcial

En el caso de dos variables, la $r$ simple tenía un significado directo: medía el grado de asociación (lineal), (mas no causal), entre la variable dependiente $Y$ y la variable explicativa $X$. Sin embargo, una vez fuera del caso de dos variables, se debe prestar cuidadosa atención a la interpretación del coeficiente de correlación simple. De (7.11.1), por ejemplo, observamos lo siguiente:

1. Aunque $r_{12} = 0$, $r_{12.3}$ no será cero a menos que $r_{13}$ o $r_{23}$, o ambos, sean cero.

2. Si $r_{12} = 0$ y $r_{13}$ y $r_{23}$ son diferentes de cero y tienen el mismo signo, $r_{12.3}$ será negativo, mientras que si son de signos opuestos, será positivo. Un ejemplo aclarará este punto. Sea $Y =$ rendimiento del cultivo, $X_2 =$ la lluvia y $X_3 =$ la temperatura. Suponga que $r_{12} = 0$, es decir, no hay asociación entre el rendimiento del cultivo y la lluvia. Tenga en cuenta, además, que $r_{13}$ es positiva y $r_{23}$ es negativa. Entonces, como lo indica (7.11.1), $r_{12.3}$ será positivo; es decir, con la temperatura constante, existe una asociación positiva entre el rendimiento del cultivo y la lluvia. Sin embargo, este resultado, en apariencia paradójico, no es sorprendente. Como la temperatura $X_3$ afecta el rendimiento $Y$ y también afecta la lluvia $X_2$, con el fin de encontrar la relación neta entre rendimiento del cultivo y lluvia se debe eliminar la influencia de la "molesta" variable temperatura. Este ejemplo muestra que el coeficiente de correlación simple puede generar resultados equivocados.

3. Los términos $r_{12.3}$ y $r_{12}$ (y comparaciones similares) no necesitan tener el mismo signo.

4. En el caso de dos variables $r^2$ se encuentra entre 0 y 1. La misma propiedad se cumple para los coeficientes de correlación parcial al cuadrado. Así, el lector debe verificar que es posible obtener la siguiente expresión a partir de (7.11.1):

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \tag{7.11.4}$$

[19] La mayoría de los programas de cómputo para análisis de regresión múltiple calculan de forma rutinaria los coeficientes de correlación simples; por tanto, los coeficientes de correlación parciales se obtienen de modo inmediato.

que da las interrelaciones entre los tres coeficientes de correlación de orden cero. Se derivan expresiones similares de las ecuaciones (7.11.2) y (7.11.3).

5. Suponga que $r_{13} = r_{23} = 0$. ¿Significa esto que $r_{12}$ también es cero? La respuesta es obvia y se desprende de (7.11.4). El hecho de que $Y$ y $X_3$ y $X_2$ y $X_3$ no estén correlacionadas no significa que $Y$ y $X_2$ no lo estén.

A propósito, observe que la expresión $r^2{}_{12.3}$ puede denominarse **coeficiente de determinación parcial** e interpretarse como la proporción de la variación en $Y$ no explicada por la variable $X_3$ que se explica por la inclusión de $X_2$ en el modelo (véase el ejercicio 7.5). Conceptualmente, es semejante a $R^2$.

Antes de continuar observe las siguientes relaciones entre $R^2$, los coeficientes de correlación simple y los coeficientes de correlación parcial:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \tag{7.11.5}$$

$$R^2 = r_{12}^2 + \left(1 - r_{12}^2\right) r_{13.2}^2 \tag{7.11.6}$$

$$R^2 = r_{13}^2 + \left(1 - r_{13}^2\right) r_{12.3}^2 \tag{7.11.7}$$

Para terminar esta sección, considere lo siguiente: Se planteó antes que $R^2$ no disminuye si se introduce una variable explicativa adicional en el modelo, lo cual se aprecia con claridad de (7.11.6). Esta ecuación afirma que la proporción de la variación en $Y$ explicada por $X_2$ y $X_3$ conjuntamente es la suma de dos partes: la parte explicada sólo por $X_2$ $(= r_{12}^2)$ y la parte no explicada por $X_2$ $(= 1 - r_{12}^2)$, por la proporción explicada por $X_3$ después de mantener constante la influencia de $X_2$. Ahora $R^2 > r_{12}^2$ siempre que $r_{13.2}^2 > 0$. En el peor de los casos, $r_{13.2}^2$ será cero, en cuyo caso $R^2 = r_{12}^2$.

## Resumen y conclusiones

1. En este capítulo se presentó el modelo más sencillo posible de regresión lineal múltiple, a saber, el modelo de regresión con tres variables. Se entiende que el término *lineal* se refiere a linealidad en los parámetros y no necesariamente en las variables.
2. Aunque un modelo de regresión con tres variables es, en muchas formas, una extensión del modelo con dos variables, hay algunos conceptos nuevos, como *coeficientes de regresión parcial, coeficientes de correlación parcial, coeficiente de correlación múltiple, $R^2$ ajustada y no ajustada (por grados de libertad), multicolinealidad* y *sesgo de especificación.*
3. En este capítulo se consideró también la forma funcional del modelo de regresión múltiple, como la *función de producción Cobb-Douglas* y el *modelo de regresión polinomial*.
4. Si bien $R^2$ y $R^2$ ajustada son medidas globales que indican la forma en que el modelo escogido se ajusta a un conjunto dado de datos, no debe exagerarse su importancia. Los aspectos críticos son las expectativas teóricas en que se basa el modelo en términos de los signos *a priori* de los coeficientes de las variables incluidas en él y, como se muestra en el siguiente capítulo, su significancia estadística.
5. Los resultados presentados en este capítulo se generalizan fácilmente a un modelo de regresión lineal múltiple que implique cualquier número de regresoras. Pero el álgebra se vuelve tediosa. Este tedio se evita al recurrir al álgebra matricial. Para el lector interesado, la extensión al modelo de regresión de $k$ variables mediante álgebra matricial se presenta en el **apéndice C**, opcional. Pero el lector general puede leer el resto del texto sin conocer mucho de álgebra matricial.

## EJERCICIOS

### Preguntas

7.1. Considere los datos de la tabla 7.5.

**TABLA 7.5**

| $Y$ | $X_2$ | $X_3$ |
|---|---|---|
| 1 | 1 | 2 |
| 3 | 2 | 1 |
| 8 | 3 | −3 |

Con base en estos datos, estime las siguientes regresiones:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \quad (1)$$

$$Y_i = \lambda_1 + \lambda_3 X_{3i} + u_{2i} \quad (2)$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (3)$$

*Nota:* Estime sólo los coeficientes y no los errores estándar.

*a*) ¿Es $\alpha_2 = \beta_2$? ¿Por qué?

*b*) ¿Es $\lambda_3 = \beta_3$? ¿Por qué?

¿Qué conclusión importante obtiene de este ejercicio?

7.2. De los siguientes datos estime los coeficientes de regresión parcial, sus errores estándar y los valores $R^2$ ajustada y sin ajustar:

$$\bar{Y} = 367.693 \qquad \bar{X}_2 = 402.760 \qquad \bar{X}_3 = 8.0$$

$$\sum(Y_i - \bar{Y})^2 = 6\,6042.269 \qquad \sum(X_{2i} - \bar{X}_2)^2 = 84\,855.096$$

$$\sum(X_{3i} - \bar{X}_3)^2 = 280.000 \qquad \sum(Y_i - \bar{Y})(X_{2i} - \bar{X}_2) = 74\,778.346$$

$$\sum(Y_i - \bar{Y})(X_{3i} - \bar{X}_3) = 4\,250.900 \qquad \sum(X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) = 4\,796.000$$

$$n = 15$$

7.3. Demuestre que la ecuación (7.4.7) se expresa también como

$$\hat{\beta}_2 = \frac{\sum y_i(x_{2i} - b_{23}x_{3i})}{\sum(x_{2i} - b_{23}x_{3i})^2}$$

$$= \frac{\text{covariación neta (de } x_3\text{) entre } y \text{ y } x_2}{\text{variación neta (de } x_3\text{) en } x_2}$$

donde $b_{23}$ es el coeficiente de la pendiente en la regresión de $X_2$ sobre $X_3$. (*Sugerencia:* Recuerde que $b_{23} = \sum x_{2i}x_{3i} / \sum x_{3i}^2$.)

7.4. En un modelo de regresión múltiple se le indica que el término de error $u_i$ tiene la siguiente distribución de probabilidades: $u_i \sim N(0, 4)$. ¿Cómo llevaría a cabo un experimento *Monte Carlo* para verificar que la verdadera varianza es de hecho 4?

7.5. Demuestre que $r_{12.3}^2 = (R^2 - r_{13}^2)/(1 - r_{13}^2)$ e interprete la ecuación.

7.6. Si la relación $\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 = 0$ se cumple para todos los valores de $X_1$, $X_2$ y $X_3$, encuentre los valores de los tres coeficientes de correlación parcial.

7.7. ¿Es posible obtener los siguientes resultados de un conjunto de datos?

*a*) $r_{23} = 0.9, r_{13} = -0.2, r_{12} = 0.8$

*b*) $r_{12} = 0.6, r_{23} = -0.9, r_{31} = -0.5$

*c*) $r_{21} = 0.01, r_{13} = 0.66, r_{23} = -0.7$

7.8. Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 \text{ Educación}_i + \beta_2 \text{ Años de experiencia} + u_i$$

Suponga que hace caso omiso de la variable correspondiente a los años de experiencia. ¿Qué tipos de problemas o de sesgos esperaría que surgieran? Explique de manera verbal.

7.9. Demuestre que $\beta_2$ y $\beta_3$ en (7.9.2) dan en realidad las elasticidades producción del trabajo y del capital. (Esta pregunta puede resolverse sin utilizar cálculo; sólo recuerde la definición del coeficiente de elasticidad y que un cambio en el logaritmo de una variable es relativo, suponiendo que los cambios sean pequeños.)

7.10. Considere el modelo de regresión lineal de tres variables analizado en este capítulo.

*a*) Suponga que se multiplican todos los valores $X_2$ por 2. ¿Cuál será el efecto de este escalamiento, si es que se produce alguno, sobre las estimaciones de los parámetros y sus errores estándar?

*b*) Ahora, en lugar de *a*), suponga que se multiplican todos los valores $Y$ por 2. ¿Cuál sería el efecto de esto, si es que hay alguno, sobre los parámetros estimados y sus errores estándar?

7.11. En general, $R^2 \neq r_{12}^2 + r_{13}^2$, pero esto sólo se cumple si $r_{23} = 0$. Comente y resalte la importancia de este hallazgo. [*Sugerencia:* Consulte la ecuación (7.11.5)].

7.12. Considere los siguientes modelos.*

$$\text{Modelo A:} \quad Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_{1t}$$

$$\text{Modelo B:} \quad (Y_t - X_{2t}) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_{2t}$$

*a*) ¿Serán iguales las estimaciones de MCO de $\alpha_1$ y $\beta_1$? ¿Por qué?

*b*) ¿Serán iguales las estimaciones de MCO de $\alpha_3$ y $\beta_3$? ¿Por qué?

*c*) ¿Cuál es la relación entre $\alpha_2$ y $\beta_2$?

*d*) ¿Es posible comparar los términos $R^2$ de los dos modelos? ¿Por qué?

7.13. Suponga que estima la función de consumo†

$$Y_i = \alpha_1 + \alpha_2 X_i + u_{1i}$$

y la función de ahorro

$$Z_i = \beta_1 + \beta_2 X_i + u_{2i}$$

donde $Y$ = consumo, $Z$ = ahorro, $X$ = ingreso y $X = Y + Z$, es decir, el ingreso es igual al consumo más el ahorro.

*a*) ¿Cuál es la relación, si existe, entre $\alpha_2$ y $\beta_2$? Muestre sus cálculos.

*b*) ¿Será la suma de los residuos al cuadrado, SCR, la misma para los dos modelos? Explique.

*c*) ¿Se pueden comparar los términos $R^2$ de los dos modelos? ¿Por qué?

7.14. Suponga que expresa el modelo Cobb-Douglas de (7.9.1) de la siguiente manera:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i$$

Si expresa este modelo en términos logarítmicos, tendrá ln $u_i$ como el término de perturbación en el lado derecho.

*a*) ¿Qué supuestos probabilísticos debe hacer sobre ln $u_i$ para aplicar el MCRLN? ¿Cómo probaría esto con los datos de la tabla 7.3?

*b*) ¿Se aplican los mismos supuestos a $u_i$? ¿Por qué?

---

* Adaptado de Wojciech W. Charemza y Derek F. Deadman, *Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, Edward Elgar, Brookfield, Vermont, 1992, p. 18.

† Adaptado de Peter Kennedy, *A Guide To Econometrics*, 3a. ed., The MIT Press, Cambridge, Massachusetts, 1992, p. 308, pregunta 9.

7.15. *Regresión a través del origen*. Considere la siguiente regresión a través del origen:

$$Y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

*a*) ¿Qué pasos seguiría al estimar las incógnitas?

*b*) ¿Será $\sum \hat{u}_i$ cero para este modelo? ¿Por qué?

*c*) ¿Será $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$ para este modelo?

*d*) ¿Cuándo utilizaría un modelo de este tipo?

*e*) ¿Puede generalizar los resultados para el modelo de *k* variables?

(*Sugerencia:* Siga el análisis para el caso de dos variables del capítulo 6.)

## Ejercicios empíricos

7.16. *La demanda de rosas*.* En la tabla 7.6 se presentan datos trimestrales sobre estas variables:

$Y$ = cantidad de rosas vendidas, docenas

$X_2$ = precio promedio al mayoreo de las rosas, \$/docena

$X_3$ = precio promedio al mayoreo de los claveles, \$/docena

$X_4$ = ingreso familiar disponible promedio semanal, \$/semana

$X_5$ = variable de tendencia que toma valores de 1, 2, y así sucesivamente, durante el periodo 1971-III a 1975-II en el área metropolitana de Detroit.

Se le pide considerar las siguientes funciones de demanda:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 X_{5t} + u_t$$

*a*) Estime los parámetros del modelo lineal e interprete los resultados.

*b*) Estime los parámetros del modelo log-lineal e interprete los resultados.

**TABLA 7.6**
**Demanda trimestral de rosas en el área metropolitana de Detroit, de 1971-III a 1975-II**

| Año y trimestre | $Y$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| 1971-III | 11 484 | 2.26 | 3.49 | 158.11 | 1 |
| -IV | 9 348 | 2.54 | 2.85 | 173.36 | 2 |
| 1972-I | 8 429 | 3.07 | 4.06 | 165.26 | 3 |
| -II | 10 079 | 2.91 | 3.64 | 172.92 | 4 |
| -III | 9 240 | 2.73 | 3.21 | 178.46 | 5 |
| -IV | 8 862 | 2.77 | 3.66 | 198.62 | 6 |
| 1973-I | 6 216 | 3.59 | 3.76 | 186.28 | 7 |
| -II | 8 253 | 3.23 | 3.49 | 188.98 | 8 |
| -III | 8 038 | 2.60 | 3.13 | 180.49 | 9 |
| -IV | 7 476 | 2.89 | 3.20 | 183.33 | 10 |
| 1974-I | 5 911 | 3.77 | 3.65 | 181.87 | 11 |
| -II | 7 950 | 3.64 | 3.60 | 185.00 | 12 |
| -III | 6 134 | 2.82 | 2.94 | 184.00 | 13 |
| -IV | 5 868 | 2.96 | 3.12 | 188.20 | 14 |
| 1975-I | 3 160 | 4.24 | 3.58 | 175.67 | 15 |
| -II | 5 872 | 3.69 | 3.53 | 188.00 | 16 |

* El autor agradece a Joe Walsh por recopilar estos datos de un mayorista en el área metropolitana de Detroit y por su subsiguiente procesamiento.

*c*) $\beta_2$, $\beta_3$ y $\beta_4$ dan respectivamente las elasticidades de la demanda respecto del *precio propio, precio cruzado* e *ingreso*. ¿Cuáles son, *a priori*, los signos de estas elasticidades? ¿Concuerdan estos resultados con las expectativas *a priori*?

*d*) ¿Cómo calcularía las elasticidades precio propio, precio cruzado e ingreso en el modelo lineal?

*e*) Con base en el análisis, ¿cuál modelo, si existe, escogería y por qué?

7.17. *Actividad de exploración.* Reciben el nombre de "pozos de exploración" los que se perforan para encontrar y producir petróleo o gas natural en una zona mejorada, o para encontrar una nueva reserva en un yacimiento donde antes se encontró petróleo o gas natural, o para extender el límite de una reserva de petróleo o gas conocida. La tabla 7.7 contiene datos sobre estas variables:*

$Y$ = número de pozos de exploración perforados

$X_2$ = precio en la cabeza del pozo en el periodo anterior (en dólares constantes, 1972 = 100)

$X_3$ = producción interna

$X_4$ = PNB en dólares constantes (1972 = 100)

$X_5$ = variable de tendencia, 1948 = 1, 1949 = 2,..., 1978 = 31

Vea si el siguiente modelo se ajusta a los datos:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

*a*) ¿Puede ofrecer una justificación *a priori* para este modelo?

*b*) Si el modelo es aceptable, estime los parámetros del modelo y sus errores estándar, y obtenga $R^2$ y $\bar{R}^2$.

*c*) Comente sus resultados desde el punto de vista de sus expectativas *a priori*.

*d*) ¿Qué otra especificación sugeriría para explicar la actividad de exploración? ¿Por qué?

7.18. *Desembolsos del presupuesto de defensa de Estados Unidos, 1962-1981*. Para explicar el presupuesto de defensa de Estados Unidos, considere el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

donde $Y_t$ = desembolsos del presupuesto de defensa durante el año $t$, \$ miles de millones

$X_{2t}$ = PNB durante el año $t$, \$ miles de millones

$X_{3t}$ = ventas militares de Estados Unidos/ayuda en el año $t$, \$ miles de millones

$X_{4t}$ = ventas de la industria aeroespacial, \$ miles de millones

$X_{5t}$ = conflictos militares que implican a más de 100 000 soldados. Esta variable adquiere el valor de 1 cuando participan 100 000 soldados o más, y es igual a cero cuando el número de soldados no llega a 100 000.

Para probar este modelo, se proporcionan datos en la tabla 7.8.

*a*) Estime los parámetros de este modelo y sus errores estándar, y obtenga $R^2$, $R^2$ modificada y $\bar{R}^2$.

*b*) Comente los resultados, considerando cualquier expectativa *a priori* que tenga sobre la relación entre $Y$ y las diversas variables $X$.

*c*) ¿Qué otra(s) variable(s) incluiría en el modelo y por qué?

* El autor agradece a Raymond Savino por recopilar y procesar estos datos.

**TABLA 7.7**
**Actividad de exploración**

Fuente: Energy Information Administration, *Informe al Congreso*, 1978.

| Miles de pozos de exploración ($Y$) | Precio por barril, \$ constantes ($X_2$) | Producción nacional (millones de barriles diarios) ($X_3$) | PNB, miles de millones de \$ constantes ($X_4$) | Tiempo ($X_5$) |
|---|---|---|---|---|
| 8.01 | 4.89 | 5.52 | 487.67 | 1948 = 1 |
| 9.06 | 4.83 | 5.05 | 490.59 | 1949 = 2 |
| 10.31 | 4.68 | 5.41 | 533.55 | 1950 = 3 |
| 11.76 | 4.42 | 6.16 | 576.57 | 1951 = 4 |
| 12.43 | 4.36 | 6.26 | 598.62 | 1952 = 5 |
| 13.31 | 4.55 | 6.34 | 621.77 | 1953 = 6 |
| 13.10 | 4.66 | 6.81 | 613.67 | 1954 = 7 |
| 14.94 | 4.54 | 7.15 | 654.80 | 1955 = 8 |
| 16.17 | 4.44 | 7.17 | 668.84 | 1956 = 9 |
| 14.71 | 4.75 | 6.71 | 681.02 | 1957 = 10 |
| 13.20 | 4.56 | 7.05 | 679.53 | 1958 = 11 |
| 13.19 | 4.29 | 7.04 | 720.53 | 1959 = 12 |
| 11.70 | 4.19 | 7.18 | 736.86 | 1960 = 13 |
| 10.99 | 4.17 | 7.33 | 755.34 | 1961 = 14 |
| 10.80 | 4.11 | 7.54 | 799.15 | 1962 = 15 |
| 10.66 | 4.04 | 7.61 | 830.70 | 1963 = 16 |
| 10.75 | 3.96 | 7.80 | 874.29 | 1964 = 17 |
| 9.47 | 3.85 | 8.30 | 925.86 | 1965 = 18 |
| 10.31 | 3.75 | 8.81 | 980.98 | 1966 = 19 |
| 8.88 | 3.69 | 8.66 | 1 007.72 | 1967 = 20 |
| 8.88 | 3.56 | 8.78 | 1 051.83 | 1968 = 21 |
| 9.70 | 3.56 | 9.18 | 1 078.76 | 1969 = 22 |
| 7.69 | 3.48 | 9.03 | 1 075.31 | 1970 = 23 |
| 6.92 | 3.53 | 9.00 | 1 107.48 | 1971 = 24 |
| 7.54 | 3.39 | 8.78 | 1 171.10 | 1972 = 25 |
| 7.47 | 3.68 | 8.38 | 1 234.97 | 1973 = 26 |
| 8.63 | 5.92 | 8.01 | 1 217.81 | 1974 = 27 |
| 9.21 | 6.03 | 7.78 | 1 202.36 | 1975 = 28 |
| 9.23 | 6.12 | 7.88 | 1 271.01 | 1976 = 29 |
| 9.96 | 6.05 | 7.88 | 1 332.67 | 1977 = 30 |
| 10.78 | 5.89 | 8.67 | 1 385.10 | 1978 = 31 |

7.19. *Demanda de carne de pollo en Estados Unidos, 1960-1982*. Para estudiar el consumo *per cápita* de carne de pollo en Estados Unidos se presentan los datos de la tabla 7.9,

donde $Y$ = consumo *per cápita* de carne de pollo, lbs

$X_2$ = ingreso *per cápita* real disponible, \$

$X_3$ = precio real al menudeo del pollo por lb, ¢

$X_4$ = precio real al menudeo del cerdo por lb, ¢

$X_5$ = precio real al menudeo de la carne de res por lb, ¢

$X_6$ = Precio real compuesto de los sustitutos de pollo por lb, ¢, el cual es un promedio ponderado de los precios reales al menudeo por libra de carne de cerdo y de res; las ponderaciones son los consumos relativos de la carne de res y de cerdo en el consumo total de estos productos.

**TABLA 7.8**
**Desembolsos del presupuesto de defensa de Estados Unidos, 1962-1981**

Fuente: Albert Lucchino recopiló estos datos de diversas publicaciones gubernamentales.

| Año | Desembolsos del presupuesto de defensa ($Y$) | PNB ($X_2$) | Ventas/asistencias militares de Estados Unidos ($X_3$) | Ventas de la industria aeroespacial ($X_4$) | Conflictos 100 000+ ($X_5$) |
|---|---|---|---|---|---|
| 1962 | 51.1 | 560.3 | 0.6 | 16.0 | 0 |
| 1963 | 52.3 | 590.5 | 0.9 | 16.4 | 0 |
| 1964 | 53.6 | 632.4 | 1.1 | 16.7 | 0 |
| 1965 | 49.6 | 684.9 | 1.4 | 17.0 | 1 |
| 1966 | 56.8 | 749.9 | 1.6 | 20.2 | 1 |
| 1967 | 70.1 | 793.9 | 1.0 | 23.4 | 1 |
| 1968 | 80.5 | 865.0 | 0.8 | 25.6 | 1 |
| 1969 | 81.2 | 931.4 | 1.5 | 24.6 | 1 |
| 1970 | 80.3 | 992.7 | 1.0 | 24.8 | 1 |
| 1971 | 77.7 | 1 077.6 | 1.5 | 21.7 | 1 |
| 1972 | 78.3 | 1 185.9 | 2.95 | 21.5 | 1 |
| 1973 | 74.5 | 1 326.4 | 4.8 | 24.3 | 0 |
| 1974 | 77.8 | 1 434.2 | 10.3 | 26.8 | 0 |
| 1975 | 85.6 | 1 549.2 | 16.0 | 29.5 | 0 |
| 1976 | 89.4 | 1 718.0 | 14.7 | 30.4 | 0 |
| 1977 | 97.5 | 1 918.3 | 8.3 | 33.3 | 0 |
| 1978 | 105.2 | 2 163.9 | 11.0 | 38.0 | 0 |
| 1979 | 117.7 | 2 417.8 | 13.0 | 46.2 | 0 |
| 1980 | 135.9 | 2 633.1 | 15.3 | 57.6 | 0 |
| 1981 | 162.1 | 2 937.7 | 18.0 | 68.9 | 0 |

**TABLA 7.9**
**Demanda de carne de pollo en Estados Unidos, 1960-1982**

Fuente: Los datos sobre $Y$ provienen de *Citibase*, y los datos sobre $X_2$ hasta $X_6$, del Departamento de Agricultura de Estados Unidos. El autor agradece a Robert J. Fisher por reunir los datos y el análisis estadístico.

| Año | $Y$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| 1960 | 27.8 | 397.5 | 42.2 | 50.7 | 78.3 | 65.8 |
| 1961 | 29.9 | 413.3 | 38.1 | 52.0 | 79.2 | 66.9 |
| 1962 | 29.8 | 439.2 | 40.3 | 54.0 | 79.2 | 67.8 |
| 1963 | 30.8 | 459.7 | 39.5 | 55.3 | 79.2 | 69.6 |
| 1964 | 31.2 | 492.9 | 37.3 | 54.7 | 77.4 | 68.7 |
| 1965 | 33.3 | 528.6 | 38.1 | 63.7 | 80.2 | 73.6 |
| 1966 | 35.6 | 560.3 | 39.3 | 69.8 | 80.4 | 76.3 |
| 1967 | 36.4 | 624.6 | 37.8 | 65.9 | 83.9 | 77.2 |
| 1968 | 36.7 | 666.4 | 38.4 | 64.5 | 85.5 | 78.1 |
| 1969 | 38.4 | 717.8 | 40.1 | 70.0 | 93.7 | 84.7 |
| 1970 | 40.4 | 768.2 | 38.6 | 73.2 | 106.1 | 93.3 |
| 1971 | 40.3 | 843.3 | 39.8 | 67.8 | 104.8 | 89.7 |
| 1972 | 41.8 | 911.6 | 39.7 | 79.1 | 114.0 | 100.7 |
| 1973 | 40.4 | 931.1 | 52.1 | 95.4 | 124.1 | 113.5 |
| 1974 | 40.7 | 1 021.5 | 48.9 | 94.2 | 127.6 | 115.3 |
| 1975 | 40.1 | 1 165.9 | 58.3 | 123.5 | 142.9 | 136.7 |
| 1976 | 42.7 | 1 349.6 | 57.9 | 129.9 | 143.6 | 139.2 |
| 1977 | 44.1 | 1 449.4 | 56.5 | 117.6 | 139.2 | 132.0 |
| 1978 | 46.7 | 1 575.5 | 63.7 | 130.9 | 165.5 | 132.1 |
| 1979 | 50.6 | 1 759.1 | 61.6 | 129.8 | 203.3 | 154.4 |
| 1980 | 50.1 | 1 994.2 | 58.9 | 128.0 | 219.6 | 174.9 |
| 1981 | 51.7 | 2 258.1 | 66.4 | 141.0 | 221.6 | 180.8 |
| 1982 | 52.9 | 2 478.7 | 70.4 | 168.2 | 232.6 | 189.4 |

*Nota:* Los precios reales se obtuvieron al dividir los precios nominales entre el Índice de Precios al Consumidor correspondiente a alimentos.

Ahora considere las siguientes funciones de demanda:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \quad (2)$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \quad (3)$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \quad (4)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \quad (5)$$

De la teoría microeconómica, se sabe que la demanda de un bien suele depender del ingreso real del consumidor, del precio real del bien y de los precios reales de los bienes complementarios o que compiten con él. Ante estas consideraciones, responda las siguientes preguntas.

*a*) Entre las funciones de demanda que aquí se dan, ¿cuál escogería y por qué?

*b*) ¿Cómo interpretaría los coeficientes de $\ln X_{2t}$ y $\ln X_{3t}$ en estos modelos?

*c*) ¿Cuál es la diferencia entre las especificaciones (2) y (4)?

*d*) ¿Qué problemas prevé si adopta la especificación (4)? (*Sugerencia:* Los precios de la carne de cerdo y de res se incluyen con el del pollo.)

*e*) Como la especificación (5) incluye el precio compuesto de la carne de res y de cerdo, ¿preferiría la función de demanda (5) a la función (4)? ¿Por qué?

*f*) ¿La carne de cerdo y la de res son productos que compiten con el pollo o que lo sustituyen? ¿Cómo sabe?

*g*) Suponga que la función (5) es la de demanda "correcta". Estime los parámetros de este modelo, obtenga sus errores estándar así como $R^2$, $\bar{R}^2$ y $R^2$ modificada. Interprete sus resultados.

*h*) Ahora suponga que corre el modelo "incorrecto" (2). Evalúe las consecuencias de esta mala especificación considerando los valores de $\gamma_2$ y $\gamma_3$ en relación con $\beta_2$ y $\beta_3$, respectivamente. (*Sugerencia:* Preste atención al análisis de la sección 7.7.)

7.20. En un estudio de rotación de empleados en el mercado laboral, James F. Ragan, Jr., obtuvo los siguientes resultados para la economía de Estados Unidos de 1950-I a 1979-IV.* (Las cifras entre paréntesis son los estadísticos *t* estimados.)

$$\begin{aligned}\widehat{\ln Y_t} = {} & \underset{(4.28)}{4.47} - \underset{(-5.31)}{0.34} \ln X_{2t} + \underset{(3.64)}{1.22} \ln X_{3t} + \underset{(3.10)}{1.22} \ln X_{4t} \\ & + \underset{(1.10)}{0.80} \ln X_{5t} - \underset{(-3.09)}{0.0055} X_{6t} \qquad \bar{R}^2 = 0.5370\end{aligned}$$

*Nota:* Estudiaremos los estadísticos *t* en el próximo capítulo.

donde $Y$ = tasa de renuncias laborales en el sector manufacturero, definida como el número de personas que deja su trabajo voluntariamente por cada 100 empleados

$X_2$ = variable instrumental o "representante" de la tasa de desempleo de hombres adultos

$X_3$ = porcentaje de empleados menores de 25 años

$X_4 = N_{t-1}/N_{t-4}$ = tasa de empleo en el sector manufacturero, en el trimestre $(t-1)$, respecto de la del trimestre $(t-4)$

$X_5$ = porcentaje de mujeres empleadas

$X_6$ = tendencia de tiempo (1950-I = 1)

* Fuente: Véase el artículo de Ragan "Turnover in the Labor Market: A Study of Quit and Layoff Rates", *Economic Review*, Federal Reserve Bank of Kansas City, mayo de 1981, pp. 13-22.

*a*) Interprete los resultados anteriores.
*b*) ¿Puede justificarse *a priori* la relación negativa observada entre los logaritmos de $Y$ y $X_2$?
*c*) ¿Por qué es positivo el coeficiente de ln $X_3$?
*d*) Como el coeficiente de tendencia es negativo, ¿de qué porcentaje es el descenso temporal en la tasa de retiro laboral, y por qué se presenta dicho descenso?
*e*) ¿Es la $\bar{R}^2$ "muy" baja?
*f*) ¿Puede estimar los errores estándar de los coeficientes de regresión a partir de los datos dados? ¿Por qué?

7.21. Considere la siguiente función de demanda de dinero para Estados Unidos durante el periodo 1980-1998:

$$M_t = \beta_1 Y_t^{\beta_2} r_t^{\beta_3} e^{u_t}$$

donde $M =$ demanda real de dinero, de acuerdo con la definición $M_2$ de dinero
$Y =$ PIB real
$r =$ tasa de interés

Para estimar la anterior función de demanda de dinero se presentan los datos de la tabla 7.10.

*Nota:* Para convertir cantidades nominales a reales, divida $M$ y PIB entre IPC. No es necesario dividir la tasa de interés variable entre el IPC. También tenga en cuenta que se proporcionaron dos tasas de interés, una de corto plazo, medida de acuerdo con la tasa de interés de los bonos del Tesoro a tres meses, y otra de largo plazo, medida según el rendimiento de los bonos del Tesoro a 30 años, según la línea de estudios empíricos previos que emplearon ambos tipos de tasas de interés.

**TABLA 7.10**
**Demanda de dinero en Estados Unidos, 1980-1998**

Fuente: *Economic Report of the President*, 2000, tablas B-1, B-58, B-67 y B-71.

| Observación | PIB | $M_2$ | IPC | TILP | TITM |
|---|---|---|---|---|---|
| 1980 | 2 795.6 | 1 600.4 | 82.4 | 11.27 | 11.506 |
| 1981 | 3 131.3 | 1 756.1 | 90.9 | 13.45 | 14.029 |
| 1982 | 3 259.2 | 1 911.2 | 96.5 | 12.76 | 10.686 |
| 1983 | 3 534.9 | 2 127.8 | 99.6 | 11.18 | 8.630 |
| 1984 | 3 932.7 | 2 311.7 | 103.9 | 12.41 | 9.580 |
| 1985 | 4 213.0 | 2 497.4 | 107.6 | 10.79 | 7.480 |
| 1986 | 4 452.9 | 2 734.0 | 109.6 | 7.78 | 5.980 |
| 1987 | 4 742.5 | 2 832.8 | 113.6 | 8.59 | 5.820 |
| 1988 | 5 108.3 | 2 995.8 | 118.3 | 8.96 | 6.690 |
| 1989 | 5 489.1 | 3 159.9 | 124.0 | 8.45 | 8.120 |
| 1990 | 5 803.2 | 3 279.1 | 130.7 | 8.61 | 7.510 |
| 1991 | 5 986.2 | 3 379.8 | 136.2 | 8.14 | 5.420 |
| 1992 | 6 318.9 | 3 434.1 | 140.3 | 7.67 | 3.450 |
| 1993 | 6 642.3 | 3 487.5 | 144.5 | 6.59 | 3.020 |
| 1994 | 7 054.3 | 3 502.2 | 148.2 | 7.37 | 4.290 |
| 1995 | 7 400.5 | 3 649.3 | 152.4 | 6.88 | 5.510 |
| 1996 | 7 813.2 | 3 824.2 | 156.9 | 6.71 | 5.020 |
| 1997 | 8 300.8 | 4 046.7 | 160.5 | 6.61 | 5.070 |
| 1998 | 8 759.9 | 4 401.4 | 163.0 | 5.58 | 4.810 |

*Notas:* PIB: producto interno bruto (miles de millones de dólares).
$M_2$: oferta de dinero $M_2$.
IPC: índice de precios al consumidor.
TILP: tasa de interés de largo plazo (bonos del Tesoro a 30 años).
TITM: tasa de interés de los bonos del Tesoro a tres meses (% anual).

*a*) Con los datos anteriores, calcule la función de demanda anterior. ¿Cuáles son las elasticidades del ingreso y de la tasa de interés de la demanda de dinero?

*b*) En lugar de estimar la función demanda anterior, suponga que debe ajustar la función $(M/Y)_t = \alpha_1 r_t^{\alpha_2} e^{u_t}$. ¿Cómo interpretaría los resultados? Muestre los cálculos necesarios.

*c*) ¿Cómo decidiría cuál es la mejor especificación? (*Nota:* Se ofrece una prueba estadística formal en el capítulo 8.)

7.22. La tabla 7.11 suministra datos sobre el sector manufacturero de la economía griega de 1961 a 1987.

*a*) Vea si la función producción de Cobb-Douglas se ajusta a los datos de la tabla e interprete los resultados. ¿A qué conclusión general llega?

*b*) Ahora considere el siguiente modelo:

$$\text{Producción/trabajo} = A(K)L)^{\beta} e^{u}$$

donde la variable regresada representa la productividad del trabajo, y la regresora, la razón capital-trabajo. ¿Cuál es la importancia económica de dicha relación, si existe alguna? Estime los parámetros de este modelo e interprete los resultados.

**TABLA 7.11**
**Sector industrial griego**

Fuente: El autor está en deuda con George K. Zestos, de la Christopher Newport University, Virginia, por estos datos.

| Observación | Producción* | Capital | Trabajo† | Razón capital-trabajo |
|---|---|---|---|---|
| 1961 | 35.858 | 59.600 | 637.0 | 0.0936 |
| 1962 | 37.504 | 64.200 | 643.2 | 0.0998 |
| 1963 | 40.378 | 68.800 | 651.0 | 0.1057 |
| 1964 | 46.147 | 75.500 | 685.7 | 0.1101 |
| 1965 | 51.047 | 84.400 | 710.7 | 0.1188 |
| 1966 | 53.871 | 91.800 | 724.3 | 0.1267 |
| 1967 | 56.834 | 99.900 | 735.2 | 0.1359 |
| 1968 | 65.439 | 109.100 | 760.3 | 0.1435 |
| 1969 | 74.939 | 120.700 | 777.6 | 0.1552 |
| 1970 | 80.976 | 132.000 | 780.8 | 0.1691 |
| 1971 | 90.802 | 146.600 | 825.8 | 0.1775 |
| 1972 | 101.955 | 162.700 | 864.1 | 0.1883 |
| 1973 | 114.367 | 180.600 | 894.2 | 0.2020 |
| 1974 | 101.823 | 197.100 | 891.2 | 0.2212 |
| 1975 | 107.572 | 209.600 | 887.5 | 0.2362 |
| 1976 | 117.600 | 221.900 | 892.3 | 0.2487 |
| 1977 | 123.224 | 232.500 | 930.1 | 0.2500 |
| 1978 | 130.971 | 243.500 | 969.9 | 0.2511 |
| 1979 | 138.842 | 257.700 | 1 006.9 | 0.2559 |
| 1980 | 135.486 | 274.400 | 1 020.9 | 0.2688 |
| 1981 | 133.441 | 289.500 | 1 017.1 | 0.2846 |
| 1982 | 130.388 | 301.900 | 1 016.1 | 0.2971 |
| 1983 | 130.615 | 314.900 | 1 008.1 | 0.3124 |
| 1984 | 132.244 | 327.700 | 985.1 | 0.3327 |
| 1985 | 137.318 | 339.400 | 977.1 | 0.3474 |
| 1986 | 137.468 | 349.492 | 1 007.2 | 0.3470 |
| 1987 | 135.750 | 358.231 | 1 000.0 | 0.3582 |

* Miles de millones de dracmas a precios constantes de 1970.
† Miles de trabajadores al año.

7.23. *Experimento Monte Carlo.* Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Le informan que $\beta_1 = 262$, $\beta_2 = -0.006$, $\beta_3 = -2.4$, $\sigma^2 = 42$ y $u_i \sim N(0, 42)$. Genere 10 conjuntos de 64 observaciones sobre $u_i$ a partir de la distribución normal dada y utilice las 64 observaciones de la tabla 6.4, donde $Y =$ MI, $X_2 =$ PIBPC, y $X_3 =$ TAM, para generar 10 conjuntos de coeficientes $\beta$ estimados (cada conjunto tendrá tres parámetros estimados). Tome los promedios de todos los coeficientes $\beta$ estimados y relaciónelos con los verdaderos valores de dichos coeficientes dados arriba. ¿A qué conclusión general llega de lo anterior?

7.24. La tabla 7.12 presenta datos del gasto de consumo real, ingreso real, riqueza real y tasas de interés reales de Estados Unidos de 1947 a 2000. Estos datos se volverán a usar en el ejercicio 8.35.

*a*) Con los datos de la tabla, estime la función de consumo lineal usando los datos de ingreso, riqueza y tasa de interés. ¿Cuál es la ecuación ajustada?

*b*) ¿Qué indican los coeficientes estimados sobre las relaciones entre las variables y el gasto de consumo?

**TABLA 7.12**
**Gasto de consumo real, ingreso real, riqueza real y tasas de interés reales de Estados Unidos, 1947-2000**

Fuentes: *C*, Yd y los índices de precios trimestrales y anuales, ajustados por la inflación (1996 = 100), Oficina de Análisis Económico, Departamento de Comercio de Estados Unidos (http://www.bea.doc.gov/bea/dn1.htm). Rendimiento nominal anual de los bonos del Tesoro a tres meses: Economic Report of the President, 2002.

Riqueza nominal = valor neto nominal a final de año de unidades familiares y organizaciones sin fines de lucro (tomado de los datos de flujo de fondos de la Reserva Federal; http://www.federalreserve.gov).

| Año | *C* | Yd | Riqueza | Tasa de interés |
|---|---|---|---|---|
| 1947 | 976.4 | 1 035.2 | 5 166.8 | −10.351 |
| 1948 | 998.1 | 1 090.0 | 5 280.8 | −4.720 |
| 1949 | 1 025.3 | 1 095.6 | 5 607.4 | 1.044 |
| 1950 | 1 090.9 | 1 192.7 | 5 759.5 | 0.407 |
| 1951 | 1 107.1 | 1 227.0 | 6 086.1 | −5.283 |
| 1952 | 1 142.4 | 1 266.8 | 6 243.9 | −0.277 |
| 1953 | 1 197.2 | 1 327.5 | 6 355.6 | 0.561 |
| 1954 | 1 221.9 | 1 344.0 | 6 797.0 | −0.138 |
| 1955 | 1 310.4 | 1 433.8 | 7 172.2 | 0.262 |
| 1956 | 1 348.8 | 1 502.3 | 7 375.2 | −0.736 |
| 1957 | 1 381.8 | 1 539.5 | 7 315.3 | −0.261 |
| 1958 | 1 393.0 | 1 553.7 | 7 870.0 | −0.575 |
| 1959 | 1 470.7 | 1 623.8 | 8 188.1 | 2.296 |
| 1960 | 1 510.8 | 1 664.8 | 8 351.8 | 1.511 |
| 1961 | 1 541.2 | 1 720.0 | 8 971.9 | 1.296 |
| 1962 | 1 617.3 | 1 803.5 | 9 091.5 | 1.396 |
| 1963 | 1 684.0 | 1 871.5 | 9 436.1 | 2.058 |
| 1964 | 1 784.8 | 2 006.9 | 10 003.4 | 2.027 |
| 1965 | 1 897.6 | 2 131.0 | 10 562.8 | 2.112 |
| 1966 | 2 006.1 | 2 244.6 | 10 522.0 | 2.020 |
| 1967 | 2 066.2 | 2 340.5 | 11 312.1 | 1.213 |
| 1968 | 2 184.2 | 2 448.2 | 12 145.4 | 1.055 |
| 1969 | 2 264.8 | 2 524.3 | 11 672.3 | 1.732 |
| 1970 | 2 314.5 | 2 630.0 | 11 650.0 | 1.166 |
| 1971 | 2 405.2 | 2 745.3 | 12 312.9 | −0.712 |
| 1972 | 2 550.5 | 2 874.3 | 13 499.9 | −0.156 |
| 1973 | 2 675.9 | 3 072.3 | 13 081.0 | 1.414 |
| 1974 | 2 653.7 | 3 051.9 | 11 868.8 | −1.043 |
| 1975 | 2 710.9 | 3 108.5 | 12 634.4 | −3.534 |
| 1976 | 2 868.9 | 3 243.5 | 13 456.8 | −0.657 |

(*continúa*)

**TABLA 7.12** (*continuación*)

| Año | C | Yd | Riqueza | Tasa de interés |
|---|---|---|---|---|
| 1977 | 2 992.1 | 3 360.7 | 13 786.3 | −1.190 |
| 1978 | 3 124.7 | 3 527.5 | 14 450.5 | 0.113 |
| 1979 | 3 203.2 | 3 628.6 | 15 340.0 | 1.704 |
| 1980 | 3 193.0 | 3 658.0 | 15 965.0 | 2.298 |
| 1981 | 3 236.0 | 3 741.1 | 15 965.0 | 4.704 |
| 1982 | 3 275.5 | 3 791.7 | 16 312.5 | 4.449 |
| 1983 | 3 454.3 | 3 906.9 | 16 944.8 | 4.691 |
| 1984 | 3 640.6 | 4 207.6 | 17 526.7 | 5.848 |
| 1985 | 3 820.9 | 4 347.8 | 19 068.3 | 4.331 |
| 1986 | 3 981.2 | 4 486.6 | 20 530.0 | 3.768 |
| 1987 | 4 113.4 | 4 582.5 | 21 235.7 | 2.819 |
| 1988 | 4 279.5 | 4 784.1 | 22 332.0 | 3.287 |
| 1989 | 4 393.7 | 4 906.5 | 23 659.8 | 4.318 |
| 1990 | 4 474.5 | 5 014.2 | 23 105.1 | 3.595 |
| 1991 | 4 466.6 | 5 033.0 | 24 050.2 | 1.803 |
| 1992 | 4 594.5 | 5 189.3 | 24 418.2 | 1.007 |
| 1993 | 4 748.9 | 5 261.3 | 25 092.3 | 0.625 |
| 1994 | 4 928.1 | 5 397.2 | 25 218.6 | 2.206 |
| 1995 | 5 075.6 | 5 539.1 | 27 439.7 | 3.333 |
| 1996 | 5 237.5 | 5 677.7 | 29 448.2 | 3.083 |
| 1997 | 5 423.9 | 5 854.5 | 32 664.1 | 3.120 |
| 1998 | 5 683.7 | 6 168.6 | 35 587.0 | 3.584 |
| 1999 | 5 968.4 | 6 320.0 | 39 591.3 | 3.245 |
| 2000 | 6 257.8 | 6 539.2 | 38 167.7 | 3.576 |

*Notas:* Año = año calendario.
$C$ = gasto de consumo real en miles de millones de dólares de 1996 ajustados por la inflación.
Yd = ingreso disponible personal real en miles de millones de dólares de 1996 ajustados por la inflación.
Riqueza = riqueza real en miles de millones de dólares de 1996 ajustados por la inflación.
Interés = rendimiento nominal anual de los bonos del Tesoro a 3 meses, ajustado por la tasa de inflación (medido por el cambio porcentual anual en el índice anual de precios, ajustado por la inflación).

La variable de riqueza nominal real se creó con datos de la medición que realiza la Junta de la Reserva Federal del valor neto a final de año de unidades familiares y organizaciones sin fines de lucro en las cuentas del flujo de fondos. El índice de precios con que se convirtió esta variable de riqueza nominal en una variable de riqueza real fue el promedio del índice de precios, ajustado por la inflación, del cuarto trimestre del año en curso y el primer trimestre del año siguiente.

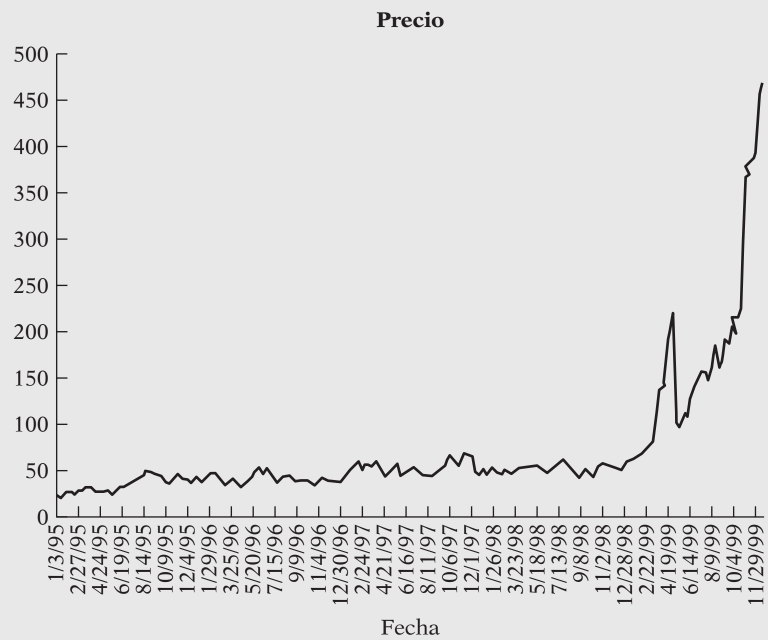
7.25. *Estimación de los precios de las acciones de Qualcomm.* Como ejemplo de la regresión polinomial considere los datos sobre los precios semanales de las acciones de Qualcomm, Inc., diseñador y fabricante de aparatos inalámbricos digitales para telecomunicaciones, de 1995 a 2000. Los datos completos se encuentran en el sitio Web del libro, en la tabla 7.13. A finales de la década de 1990, las acciones de las empresas tecnológicas fueron especialmente redituables, pero, ¿qué tipo de modelo de regresión se ajusta mejor a estos datos? La figura 7.4 muestra una gráfica elemental de los datos correspondiente a esos años.

Esta gráfica no se parece a la curva en forma de S alargada; parece haber un ligero aumento en el precio promedio de las acciones, pero luego la tasa se incrementa de manera drástica hacia el extremo derecho. A medida que la demanda de teléfonos más especializados aumentaba a pasos agigantados y el auge tecnológico seguía su curso, el precio de las acciones siguió la tendencia y aumentó a un ritmo mucho más rápido.

*a*) Estime un modelo lineal para pronosticar el *precio de cierre de las acciones* con base en el *tiempo.* ¿Le parece que este modelo se ajusta bien a los datos?

*b*) Ahora estime un modelo cuadrático y use tanto el *tiempo* como el *tiempo elevado al cuadrado.* ¿Obtuvo un mejor ajuste que en *a*)?

**FIGURA 7.4**
Precios de las acciones de Qualcomm a través del tiempo.




*c*) Por último, ajuste el siguiente *polinomio de tercer grado* o cúbico:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

donde $Y =$ precio de las acciones y $X =$ tiempo. ¿Qué modelo parece el mejor estimador de los precios de las acciones?

# Apéndice 7A

## 7A.1 Derivación de los estimadores de MCO dados en las ecuaciones (7.4.3) a (7.4.5)

Al diferenciar parcialmente la ecuación

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \tag{7.4.2}$$

respecto de las tres incógnitas e igualar a cero las ecuaciones resultantes, obtenemos

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-1) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0$$

Simplificamos lo anterior y obtenemos las ecuaciones (7.4.3) a (7.4.5).

A propósito, observe que las tres ecuaciones anteriores también se escriben como

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = 0 \qquad \text{(¿Por qué?)}$$

$$\sum \hat{u}_i X_{3i} = 0$$

lo cual muestra que las propiedades de los mínimos cuadrados se ajustan, es decir, que los residuos suman cero y que no están correlacionados con las variables explicativas $X_2$ y $X_3$.

Por cierto, observe que para obtener los estimadores de MCO del modelo de regresión lineal de $k$ variables (7.4.20) se procede análogamente. Así, primero escribimos

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

Diferenciamos esta expresión parcialmente respecto de cada incógnitas $k$, igualamos a cero las ecuaciones resultantes y reorganizamos para obtener las siguientes $k$ ecuaciones normales con $k$ incógnitas:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i} X_{ki}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i} X_{ki}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum Y_i X_{ki} = \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{2i} X_{ki} + \hat{\beta}_3 \sum X_{3i} X_{ki} + \cdots + \hat{\beta}_k \sum X_{ki}^2$$

O, si cambiamos a letras minúsculas, estas ecuaciones se expresan como

$$\sum y_i x_{2i} = \hat{\beta}_2 \sum x_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} + \cdots + \hat{\beta}_k \sum x_{2i} x_{ki}$$

$$\sum y_i x_{3i} = \hat{\beta}_2 \sum x_{2i} x_{3i} + \hat{\beta}_3 \sum x_{3i}^2 + \cdots + \hat{\beta}_k \sum x_{3i} x_{ki}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum y_i x_{ki} = \hat{\beta}_2 \sum x_{2i} x_{ki} + \hat{\beta}_3 \sum x_{3i} x_{ki} + \cdots + \hat{\beta}_k \sum x_{ki}^2$$

Debe observar, además, que el modelo de $k$ variables satisface también estas ecuaciones:

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = \cdots = \sum \hat{u}_i X_{ki} = 0$$

## 7A.2 Igualdad entre los coeficientes del PIBPC en las ecuaciones (7.3.5) y (7.6.2)

Sea $Y = \text{MI}$, $X_2 = \text{PIBPC}$ y $X_3 = \text{TAM}$, y al utilizar la forma de desviación, se tiene

$$y_i = b_{13}x_{3i} + \hat{u}_{1i} \tag{1}$$

$$x_{2i} = b_{23}x_{3i} + \hat{u}_{2i} \tag{2}$$

Ahora efectuamos la regresión en $\hat{u}_1$ sobre $\hat{u}_2$ para obtener:

$$a_1 = \frac{\sum \hat{u}_{1i}\hat{u}_{2i}}{\hat{u}_{2i}^2} = -0.0056 \qquad \text{(para este ejemplo)} \tag{3}$$

Observe que, en vista de que las $\hat{u}$ son residuos, sus valores medios son cero. Con (1) y (2), (3) se expresa como

$$a_1 = \frac{\sum(y_i - b_{13}x_{3i})(x_{2i} - b_{23}x_{3i})}{\sum(x_{2i} - b_{23}x_{3i})^2} \tag{4}$$

Expanda la expresión anterior y observe que

$$b_{23} = \frac{\sum x_{2i}x_{3i}}{\sum x_{3i}^2} \tag{5}$$

y

$$b_{13} = \frac{\sum y_i x_{3i}}{\sum x_{3i}^2} \tag{6}$$

Al hacer esas sustituciones en (4), obtenemos

$$\begin{aligned}\hat{\beta}_2 &= \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2} \\ &= -0.0056 \qquad \text{(para este ejemplo)}\end{aligned} \tag{7.4.7}$$

## 7A.3 Derivación de la ecuación (7.4.19)

Recuerde que

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

que también se escribe como

$$\hat{u}_i = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$$

donde las letras minúsculas, como es usual, indican desviaciones respecto de los valores de la media.

Ahora,

$$\begin{aligned}\sum \hat{u}_i^2 &= \sum(\hat{u}_i\hat{u}_i) \\ &= \sum \hat{u}_i(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) \\ &= \sum \hat{u}_i y_i\end{aligned}$$

donde se aprovecha que $\sum \hat{u}_i x_{2i} = \sum \hat{u}_i x_{3i} = 0$. (¿Por qué?) También

$$\sum \hat{u}_i y_i = \sum y_i \hat{u}_i = \sum y_i (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})$$

es decir,

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

que es el resultado requerido.

## 7A.4 Estimación de máxima verosimilitud del modelo de regresión múltiple

Con la extensión de las ideas presentadas en el capítulo 4, apéndice 4A, escribimos la función log de verosimilitud para el modelo de regresión lineal con $k$ variables (7.4.20) como

$$\ln L = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln (2\pi) - \frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2}{\sigma^2}$$

Diferenciamos esta función parcialmente respecto de $\beta_1, \beta_2, \ldots, \beta_k$ y $\sigma^2$, y obtenemos las siguientes $(K+1)$ ecuaciones:

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-1) \tag{1}$$

$$\frac{\partial \ln L}{\partial \beta_2} = -\frac{1}{\sigma^2} \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{2i}) \tag{2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial \ln L}{\partial \beta_k} = -\frac{1}{\sigma^2} \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{ki}) \tag{K}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2 \tag{K+1}$$

Al igualar estas ecuaciones a cero (la condición de primer orden para optimización) y definir $\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_k$ y $\tilde{\sigma}^2$ como estimadores de MV, obtenemos, después de simples manipulaciones algebraicas,

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_{2i} + \cdots + \tilde{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \tilde{\beta}_1 \sum X_{2i} + \tilde{\beta}_2 \sum X_{2i}^2 + \cdots + \tilde{\beta}_k \sum X_{2i} X_{ki}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum Y_i X_{ki} = \tilde{\beta}_1 \sum X_{ki} + \tilde{\beta}_2 \sum X_{2i} X_{ki} + \cdots + \tilde{\beta}_k \sum X_{ki}^2$$

que son precisamente las ecuaciones normales de la teoría de mínimos cuadrados, como se ve en el apéndice 7A, sección 7A.1. Por consiguiente, los estimadores de MV, los $\tilde{\beta}$, son los mismos que los estimadores de MCO, los $\hat{\beta}$, dados antes. Pero, como mencionamos en el capítulo 4, apéndice 4A, esta igualdad no es accidental.

Sustituimos los estimadores de MV (= MCO) en la ecuación número $(K+1)$ recién dada y obtenemos, después de simplificar, el siguiente estimador de MV de $\sigma^2$

$$\tilde{\sigma}^2 = \frac{1}{n} \sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} - \cdots - \tilde{\beta}_k X_{ki})^2$$

$$= \frac{1}{n} \sum \hat{u}_i^2$$

Como se anotó en el texto, este estimador difiere del estimador de MCO $\hat{\sigma}^2 = \sum \hat{u}_i^2)(n-k)$. Y, como el último es un estimador insesgado de $\sigma^2$, esta conclusión implica que el estimador de MV $\tilde{\sigma}^2$ es sesgado. Pero, como se verifica fácilmente, asintóticamente, $\tilde{\sigma}^2$ es también insesgado.

## 7A.5 Listado de EViews de la función de producción Cobb Douglas de la ecuación (7.9.4)

Variable dependiente: Y1
Método: Mínimos cuadrados
Observaciones incluidas: 51

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | 3.887600 | 0.396228 | 9.811514 | 0.0000 |
| Y2 | 0.468332 | 0.098926 | 4.734170 | 0.0000 |
| Y3 | 0.521279 | 0.096887 | 5.380274 | 0.0000 |

| | | | |
|---|---|---|---|
| R al cuadrado | 0.964175 | Media de la var. dependiente | 16.94139 |
| R al cuadrado ajustada | 0.962683 | E.E. var. dependiente | 1.380870 |
| E.E. de la regresión | 0.266752 | Criterio de inform. Akaike | 0.252028 |
| Suma de cuadrados de residuos | 3.415520 | Criterio Schwarz | 0.365665 |
| Log verosimilitud | −3.426721 | Criterio Hannan-Quinn | 0.295452 |
| Estadístico *F* | 645.9311 | Est. Durbin-Watson | 1.946387 |
| Prob. (estadístico *F*) | 0.000000 | | |

**Covarianza de estimados**

| | C | Y2 | Y3 |
|---|---|---|---|
| C | 0.156997 | 0.010364 | −0.020014 |
| Y2 | 0.010364 | 0.009786 | −0.009205 |
| Y3 | −0.020014 | −0.009205 | 0.009387 |

| Y | X2 | X3 | Y1 | Y2 | Y3 | Y1SOMBRERO | Y1RESID |
|---|---|---|---|---|---|---|---|
| 38 372 840 | 424 471 | 2 689 076 | 17.4629 | 12.9586 | 14.8047 | 17.6739 | –0.2110 |
| 1 805 427 | 19 895 | 57 997 | 14.4063 | 9.8982 | 10.9681 | 14.2407 | 0.1656 |
| 23 736 129 | 206 893 | 2 308 272 | 16.9825 | 12.2400 | 14.6520 | 17.2577 | –0.2752 |
| 26 981 983 | 304 055 | 1 376 235 | 17.1107 | 12.6250 | 14.1349 | 17.1685 | –0.0578 |
| 217 546 032 | 1 809 756 | 13 554 116 | 19.1979 | 14.4087 | 16.4222 | 19.1962 | 0.0017 |
| 19 462 751 | 180 366 | 1 790 751 | 16.7840 | 12.1027 | 14.3981 | 17.0612 | –0.2771 |
| 28 972 772 | 224 267 | 1 210 229 | 17.1819 | 12.3206 | 14.0063 | 16.9589 | 0.2229 |
| 14 313 157 | 54 455 | 421 064 | 16.4767 | 10.9051 | 12.9505 | 15.7457 | 0.7310 |
| 159 921 | 2 029 | 7 188 | 11.9824 | 7.6153 | 8.8802 | 12.0831 | –0.1007 |
| 47 289 846 | 471 211 | 2 761 281 | 17.6718 | 13.0631 | 14.8312 | 17.7366 | –0.0648 |
| 63 015 125 | 659 379 | 3 540 475 | 17.9589 | 13.3991 | 15.0798 | 18.0236 | –0.0647 |
| 1 809 052 | 17 528 | 146 371 | 14.4083 | 9.7716 | 11.8939 | 14.6640 | –0.2557 |
| 10 511 786 | 75 414 | 848 220 | 16.1680 | 11.2307 | 13.6509 | 16.2632 | –0.0952 |
| 105 324 866 | 963 156 | 5 870 409 | 18.4726 | 13.7780 | 15.5854 | 18.4646 | 0.0079 |
| 90 120 459 | 835 083 | 5 832 503 | 18.3167 | 13.6353 | 15.5790 | 18.3944 | –0.0778 |
| 39 079 550 | 336 159 | 1 795 976 | 17.4811 | 12.7253 | 14.4011 | 17.3543 | 0.1269 |
| 22 826 760 | 246 144 | 1 595 118 | 16.9434 | 12.4137 | 14.2825 | 17.1465 | –0.2030 |
| 38 686 340 | 384 484 | 2 503 693 | 17.4710 | 12.8597 | 14.7333 | 17.5903 | –0.1193 |
| 69 910 555 | 216 149 | 4 726 625 | 18.0627 | 12.2837 | 15.3687 | 17.6519 | 0.4109 |
| 7 856 947 | 82 021 | 415 131 | 15.8769 | 11.3147 | 12.9363 | 15.9301 | –0.0532 |
| 21 352 966 | 174 855 | 1 729 116 | 16.8767 | 12.0717 | 14.3631 | 17.0284 | –0.1517 |
| 46 044 292 | 355 701 | 2 706 065 | 17.6451 | 12.7818 | 14.8110 | 17.5944 | 0.0507 |
| 92 335 528 | 943 298 | 5 294 356 | 18.3409 | 13.7571 | 15.4822 | 18.4010 | –0.0601 |
| 48 304 274 | 456 553 | 2 833 525 | 17.6930 | 13.0315 | 14.8570 | 17.7353 | –0.0423 |
| 17 207 903 | 267 806 | 1 212 281 | 16.6609 | 12.4980 | 14.0080 | 17.0429 | –0.3820 |
| 47 340 157 | 439 427 | 2 404 122 | 17.6729 | 12.9932 | 14.6927 | 17.6317 | 0.0411 |

(*continúa*)

(*continuación*)

| Y | X2 | X3 | Y1 | Y2 | Y3 | Y1SOMBRERO | Y1RESID |
|---|---|---|---|---|---|---|---|
| 2 644 567 | 24 167 | 334 008 | 14.7880 | 10.0927 | 12.7189 | 15.2445 | –0.4564 |
| 14 650 080 | 163 637 | 627 806 | 16.5000 | 12.0054 | 13.3500 | 16.4692 | 0.0308 |
| 7 290 360 | 59 737 | 522 335 | 15.8021 | 10.9977 | 13.1661 | 15.9014 | –0.0993 |
| 9 188 322 | 96 106 | 507 488 | 16.0334 | 11.4732 | 13.1372 | 16.1090 | –0.0756 |
| 51 298 516 | 407 076 | 3 295 056 | 17.7532 | 12.9168 | 15.0079 | 17.7603 | –0.0071 |
| 20 401 410 | 43 079 | 404 749 | 16.8311 | 10.6708 | 12.9110 | 15.6153 | 1.2158 |
| 87 756 129 | 727 177 | 4 260 353 | 18.2901 | 13.4969 | 15.2649 | 18.1659 | 0.1242 |
| 101 268 432 | 820 013 | 4 086 558 | 18.4333 | 13.6171 | 15.2232 | 18.2005 | 0.2328 |
| 3 556 025 | 34 723 | 184 700 | 15.0842 | 10.4552 | 12.1265 | 15.1054 | –0.0212 |
| 124 986 166 | 1 174 540 | 6 301 421 | 18.6437 | 13.9764 | 15.6563 | 18.5945 | 0.0492 |
| 20 451 196 | 201 284 | 1 327 353 | 16.8336 | 12.2125 | 14.0987 | 16.9564 | –0.1229 |
| 34 808 109 | 257 820 | 1 456 683 | 17.3654 | 12.4600 | 14.1917 | 17.1208 | 0.2445 |
| 104 858 322 | 944 998 | 5 896 392 | 18.4681 | 13.7589 | 15.5899 | 18.4580 | 0.0101 |
| 6 541 356 | 68 987 | 297 618 | 15.6937 | 11.1417 | 12.6036 | 15.6756 | 0.0181 |
| 37 668 126 | 400 317 | 2 500 071 | 17.4443 | 12.9000 | 14.7318 | 17.6085 | –0.1642 |
| 4 988 905 | 56 524 | 311 251 | 15.4227 | 10.9424 | 12.6484 | 15.6056 | –0.1829 |
| 62 828 100 | 582 241 | 4 126 465 | 17.9559 | 13.2746 | 15.2329 | 18.0451 | –0.0892 |
| 172 960 157 | 1 120 382 | 11 588 283 | 18.9686 | 13.9292 | 16.2655 | 18.8899 | 0.0786 |
| 15 702 637 | 150 030 | 762 671 | 16.5693 | 11.9186 | 13.5446 | 16.5300 | 0.0394 |
| 5 418 786 | 48 134 | 276 293 | 15.5054 | 10.7817 | 12.5292 | 15.4683 | 0.0371 |
| 49 166 991 | 425 346 | 2 731 669 | 17.7107 | 12.9607 | 14.8204 | 17.6831 | 0.0277 |
| 46 164 427 | 313 279 | 1 945 860 | 17.6477 | 12.6548 | 14.4812 | 17.3630 | 0.2847 |
| 9 185 967 | 89 639 | 685 587 | 16.0332 | 11.4035 | 13.4380 | 16.2332 | –0.2000 |
| 66 964 978 | 694 628 | 3 902 823 | 18.0197 | 13.4511 | 15.1772 | 18.0988 | –0.0791 |
| 2 979 475 | 15 221 | 361 536 | 14.9073 | 9.6304 | 12.7981 | 15.0692 | –0.1620 |

*Notas:* Y1 = ln Y; Y2 = ln X2; Y3 = ln X3.
Los valores propios son 3.7861 y 187 5269, que se usarán en el capítulo 10.

# Capítulo 8

# Análisis de regresión múltiple: el problema de la inferencia

En este capítulo, continuación del 5, se amplían las ideas desarrolladas sobre estimación por intervalos y pruebas de hipótesis a modelos con tres o más variables. Aunque en muchas formas los conceptos del capítulo 5 se aplican directamente al modelo de regresión múltiple, estos modelos poseen algunas características adicionales únicas y, por tanto, recibirán más atención.

## 8.1 Una vez más, el supuesto de normalidad

Como ya sabemos, si el único objetivo es la estimación puntual de los parámetros de los modelos de regresión, basta el método de mínimos cuadrados ordinarios (MCO), que no hace supuestos sobre la distribución de probabilidad de las perturbaciones $u_i$. Sin embargo, si el objetivo no sólo es la estimación sino además la inferencia, entonces, como vimos en los capítulos 4 y 5, debemos suponer que las $u_i$ siguen alguna distribución de probabilidad.

Por las razones ya expresadas, supusimos que las $u_i$ seguían la distribución normal con media cero y varianza constante $\sigma^2$. Se mantiene el mismo supuesto para los modelos de regresión múltiple. Con el supuesto de normalidad y el análisis de los capítulos 4 y 7 se halla que los estimadores de MCO de los coeficientes de regresión parcial, idénticos a los estimadores de máxima verosimilitud (MV), son los mejores estimadores lineales insesgados (MELI).[1] Además, los estimadores $\hat{\beta}_2$, $\hat{\beta}_3$ y $\hat{\beta}_1$ están, ellos mismos, normalmente distribuidos con medias iguales a los verdaderos $\beta_2$, $\beta_3$ y $\beta_1$, y con las varianzas dadas en el capítulo 7. Además, $(n-3)\hat{\sigma}^2/\sigma^2$ sigue la distribución $\chi^2$ con $n-3$ gl, y los tres estimadores de MCO están distribuidos independientemente de $\hat{\sigma}^2$. Las pruebas son similares a las del caso de dos variables estudiado en el apéndice 3A, sección 3A. Como resultado y a partir del capítulo 5, se puede demostrar que, al

[1] Con el supuesto de normalidad, los estimadores de MCO $\hat{\beta}_2$, $\hat{\beta}_3$ y $\hat{\beta}_1$ son de varianza mínima en toda la clase de estimadores insesgados, sean lineales o no. En resumen, son MEI (mejores estimadores insesgados). Véase C.R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Nueva York, 1965, p. 258.

reemplazar $\sigma^2$ por su estimador insesgado $\hat{\sigma}^2$ en el cálculo de los errores estándar, cada una de las siguientes variables

$$t = \frac{\hat{\beta}_1 - \beta_1}{\text{ee}\,(\hat{\beta}_1)} \quad \textbf{(8.1.1)}$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{ee}\,(\hat{\beta}_2)} \quad \textbf{(8.1.2)}$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{\text{ee}\,(\hat{\beta}_3)} \quad \textbf{(8.1.3)}$$

sigue la distribución $t$ con $n - 3$ gl.

Observe que los gl son ahora $n - 3$ porque, al calcular $\sum \hat{u}_i^2$ y, por consiguiente, $\hat{\sigma}^2$, se necesita primero estimar los tres coeficientes de regresión parcial, lo cual impone por tanto tres restricciones sobre la suma de cuadrados residual (SCR) (según esta lógica, en el caso de cuatro variables habrá $n - 4$ gl, y así sucesivamente). Por consiguiente, la distribución $t$ sirve para establecer intervalos de confianza y para probar hipótesis estadísticas sobre los verdaderos coeficientes de regresión parcial poblacionales. De modo similar, con la distribución $\chi^2$ se prueban hipótesis sobre el verdadero $\sigma^2$. Para demostrar el mecanismo real utilizaremos el siguiente ejemplo ilustrativo.

**EJEMPLO 8.1**
*De nuevo, el ejemplo de la mortalidad infantil*

En el capítulo 7 efectuamos la regresión de la mortalidad infantil (MI) sobre el PIB *per cápita* (PIBPC) y la tasa de alfabetización de las mujeres (TAM) para una muestra de 64 países. Los resultados de la regresión de (7.6.2) se reproducen a continuación, con información adicional:

$$\begin{aligned}
\widehat{\text{MI}}_i &= 263.6416 \quad - \quad 0.0056\ \text{PIBPC}_i \quad - \quad 2.2316\ \text{TAM}_i \\
\text{ee} &= (11.5932) \qquad (0.0019) \qquad\qquad (0.2099) \\
t &= (22.7411) \qquad (-2.8187) \qquad (-10.6293) \qquad\qquad \textbf{(8.1.4)} \\
\text{valor } p &= (0.0000)^* \qquad (0.0065) \qquad\qquad (0.0000)^* \\
& \qquad\qquad R^2 = 0.7077 \qquad \bar{R}^2 = 0.6981
\end{aligned}$$

donde * denota un valor extremadamente bajo.

En la ecuación (8.1.4) seguimos el formato que se presentó en la ecuación (5.11.1), donde las cifras en el primer conjunto de paréntesis son los errores estándar estimados, las del segundo conjunto son los valores $t$ según la hipótesis nula de que el coeficiente de la población relevante tiene un valor de cero, y los del tercer conjunto son los valores $p$ estimados. También se dan los valores $R^2$ y $R^2$ ajustada. Ya interpretamos esta regresión en el ejemplo 7.1.

¿Y la significancia estadística de los resultados observados? Considere por ejemplo el coeficiente del PIBPC (−0.0056). ¿Es estadísticamente significativo este coeficiente, es decir, es estadísticamente diferente de cero? Asimismo, ¿es estadísticamente significativo el coeficiente de la TAM de −2.2316? ¿Ambos coeficientes son estadísticamente significativos? Para responder ésta y otras preguntas relacionadas, primero consideremos las clases de pruebas de hipótesis que se pueden encontrar en el contexto del modelo de regresión múltiple.

## 8.2 Pruebas de hipótesis en regresión múltiple: comentarios generales

Una vez fuera del mundo simple del modelo de regresión lineal con dos variables, las pruebas de hipótesis adquieren diversas e interesantes formas, como las siguientes:

1. Pruebas de hipótesis sobre un coeficiente de regresión parcial individual (sección 8.3).
2. Pruebas de significancia global del modelo de regresión múltiple estimado, es decir, ver si todos los coeficientes de pendiente parciales son iguales a cero al mismo tiempo (sección 8.4).

3. Pruebas de que dos o más coeficientes son iguales a otro (sección 8.5).
4. Pruebas de que los coeficientes de regresión parcial satisfacen ciertas restricciones (sección 8.6).
5. Pruebas de la estabilidad del modelo de regresión estimado a través del tiempo o en diferentes unidades de corte transversal (sección 8.7).
6. Pruebas sobre la forma funcional de los modelos de regresión (sección 8.8).

Como el uso de este tipo de pruebas es tan frecuente en el análisis empírico, dedicamos una sección a cada tipo.

## 8.3 Pruebas de hipótesis sobre coeficientes de regresión individuales

Con el supuesto que $u_i \sim N(0, \sigma^2)$, entonces, como se observó en la sección 8.1, podemos utilizar la prueba $t$ para demostrar una hipótesis sobre cualquier coeficiente de regresión parcial *individual*. Para ilustrar el procedimiento considere la regresión sobre la mortalidad infantil (8.1.4). Postulemos que

$$H_0: \beta_2 = 0 \qquad \text{y} \qquad H_1: \beta_2 \neq 0$$

La hipótesis nula establece que, al mantener constante $X_3$ (la tasa de alfabetización de las mujeres), $X_2$ (PIBPC) no tiene influencia (lineal) sobre $Y$ (la mortalidad infantil).[2] Para probar la hipótesis nula se utiliza la prueba $t$ dada en (8.1.2). Según el capítulo 5, si el valor de $t$ calculado excede el valor de $t$ crítico en el nivel de significancia escogido, se rechaza la hipótesis nula; de lo contrario, no se puede rechazar. Para el ejemplo ilustrativo, con (8.1.2) y la advertencia de que $\beta_2 = 0$ con la hipótesis nula, tenemos

$$t = \frac{-0.0056}{0.0020} = -2.8187 \tag{8.3.1}$$

como se muestra en la ecuación (8.1.4).

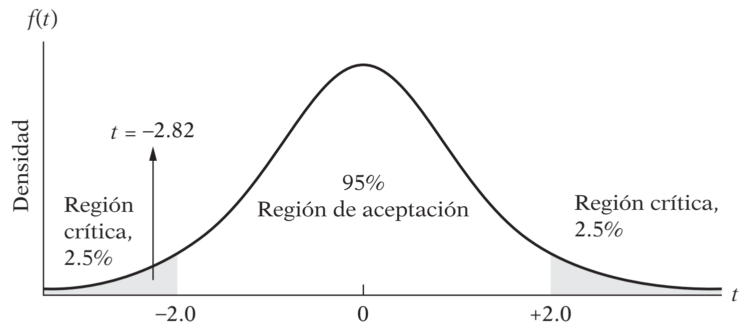
Note que tenemos 64 observaciones. Por tanto, los grados de libertad en este ejemplo son 61 (¿por qué?). Si se consulta la tabla $t$, en el **apéndice D**, no contamos con los datos correspondientes a los 61 gl. Lo más cercano que tenemos es para 60 gl. Si utilizamos esa información y suponemos $\alpha$, el nivel de significancia (es decir, la probabilidad de cometer un error de tipo I) de 5%, el valor crítico $t$ es 2.0 para una prueba de dos colas (busque $t_{\alpha/2}$ para 60 gl), o 1.671 para la prueba de una cola (busque $t_\alpha$ para 60 gl).

Para este ejemplo, la hipótesis alterna es bilateral. Por consiguiente, utilizamos el valor $t$ de dos colas. Como el valor $t$ calculado de 2.8187 (en términos absolutos) excede el valor crítico $t$ de 2, rechazamos la hipótesis nula de que el PIBPC no tiene ningún efecto sobre la mortalidad infantil. En términos más positivos, si se mantiene constante la tasa de alfabetización de las mujeres, el PIB *per cápita* tiene un efecto significativo (negativo) sobre la mortalidad infantil, como se esperaría *a priori*. De forma gráfica, la situación es la de la figura 8.1.

En la práctica, no se tiene que suponer un valor particular de $\alpha$ para llevar a cabo la prueba de hipótesis. Tan sólo se emplea el valor $p$ dado en (8.1.4), que en el caso actual es de 0.0065. La interpretación de este valor $p$ (es decir, el nivel exacto de significancia) es que si la hipótesis nula fuese verdadera, la probabilidad de obtener un valor $t$ igual a 2.8187 o mayor (en términos absolutos) sería de sólo 0.0065 o 0.65%, que de hecho es una probabilidad pequeña, mucho menor que el valor artificialmente adoptado de $\alpha = 5\%$.

[2] En la mayoría de las investigaciones empíricas, la hipótesis nula se plantea de esta forma, es decir, tomando la posición extrema (una especie de espantapájaros) de que no hay relación entre la variable dependiente y la variable explicativa en consideración. La idea aquí, para empezar, es encontrar si la relación entre las dos es trivial.

**FIGURA 8.1**
Intervalo de confianza a 95% para $t$ (60 gl).




Este ejemplo es una oportunidad para decidir si deseamos utilizar la prueba $t$ de una o dos colas. Como, *a priori*, se espera que la mortalidad infantil y el PIB *per cápita* estén negativamente relacionados (¿por qué?), se debe utilizar la prueba de una cola. Es decir, las hipótesis nula y alterna deben ser:

$$H_0: \beta_2 < 0 \qquad \text{y} \qquad H_1: \beta_2 \geq 0$$

Como el lector ya sabe, podemos rechazar la hipótesis nula con base en la prueba $t$ de una cola, en este caso. Si rechazamos la hipótesis nula en una prueba bilateral, contaremos con pruebas suficientes para rechazar el escenario unilateral siempre que el estadístico esté en la misma dirección que la prueba.

En el capítulo 5 se observó una conexión muy estrecha entre las pruebas de hipótesis y la estimación por intervalos de confianza. Para este ejemplo, el intervalo a 95% de confianza para $\beta_2$ es

$$\hat{\beta}_2 - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2)$$

que para este ejemplo se convierte en

$$-0.0056 - 2(0.0020) \leq \beta_2 \leq -0.0056 + 2(0.0020)$$

es decir,

$$-0.0096 \leq \beta_2 \leq -0.0016 \tag{8.3.2}$$

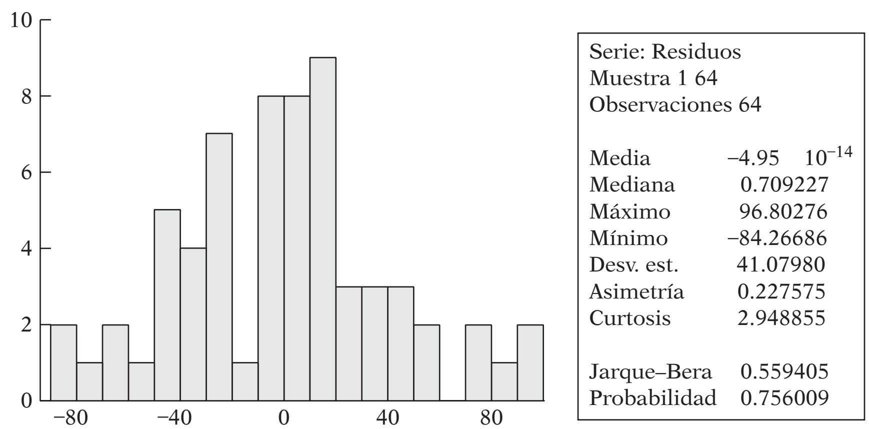
o sea, el intervalo de $-0.0096$ a $-0.0016$ incluye al verdadero coeficiente $\beta_2$, con un coeficiente de confianza de 95%. Por tanto, si se seleccionan 100 muestras de tamaño 64 y se forman 100 intervalos de confianza como el (8.3.2), esperamos que 95 de ellos contengan al verdadero parámetro de población $\beta_2$. Como el intervalo (8.3.2) no incluye el valor cero de la hipótesis nula, rechazamos tal hipótesis (que el verdadero $\beta_2$ es cero con 95% de confianza).

Por consiguiente, si se utiliza la prueba $t$ de significancia como en (8.3.1) o la estimación por intervalos de confianza como en (8.3.2), se llega a la misma conclusión. No obstante, esto no debe sorprender en vista de la estrecha relación entre la estimación por intervalos de confianza y las pruebas de hipótesis.

Según el procedimiento recién descrito se prueba la hipótesis respecto de otros parámetros del modelo de regresión para la mortalidad infantil. Los datos necesarios ya se proporcionaron en la ecuación (8.1.4). Por ejemplo, suponga que deseamos probar la hipótesis de que la tasa de alfabetización de las mujeres, si se mantiene constante la influencia del PIBPC, no tiene efecto alguno sobre la mortalidad infantil. Podemos rechazar con confianza esta hipótesis, pues, según esta hipótesis nula el valor $p$, al obtener un valor $t$ absoluto igual o mayor que 10.6, es prácticamente cero.

Antes de continuar, recuerde que el procedimiento de la prueba $t$ se basa en el supuesto de que el término de error $u_i$ sigue una distribución normal. Aunque $u_i$ no se puede observar de manera

**FIGURA 8.2**
Histograma de los residuos de la regresión (8.1.4).




directa, se observa su representante, $\hat{u}_i$, es decir, los residuos. Para la regresión sobre la mortalidad, el histograma de los residuos se muestra en la figura 8.2.

A partir del histograma, parece que los residuos están normalmente distribuidos. También podemos calcular la prueba **Jarque-Bera** (JB) de normalidad, como se muestra en la ecuación (5.12.1). En este caso, el valor JB es 0.5594, con un valor $p$ de 0.76.[3] Por tanto, al parecer, el término de error en este ejemplo sigue la distribución normal. Por supuesto, se debe tener en cuenta que la prueba JB es para muestras grandes, y que la muestra de 64 observaciones pueda no ser necesariamente grande.

## 8.4 Prueba de significancia general de la regresión muestral

En la sección anterior abordamos la prueba de significancia de los coeficientes de regresión parcial estimados individualmente, es decir, según la hipótesis separada de que cada verdadero coeficiente de regresión parcial de la población era cero. Pero ahora considere la siguiente hipótesis:

$$H_0: \beta_2 = \beta_3 = 0 \tag{8.4.1}$$

Esta hipótesis nula es una hipótesis conjunta de que $\beta_2$ y $\beta_3$ son iguales a cero en forma conjunta o simultánea. Una prueba de tal hipótesis se denomina prueba de **significancia general** de la línea de regresión observada o estimada, es decir, si $Y$ está relacionada o no linealmente con $X_2$ y $X_3$ a la vez.

¿Es demostrable la hipótesis conjunta en (8.4.1) al probar la significancia de $\hat{\beta}_2$ y $\hat{\beta}_3$ individualmente, como en la sección 8.3? La respuesta es no, y el razonamiento es el siguiente:

Al probar la significancia individual de un coeficiente de regresión parcial observado en la sección 8.3, supusimos implícitamente que cada prueba de significancia se basaba en una muestra diferente (es decir, independiente). Así, en la prueba de significancia de $\hat{\beta}_2$ según la hipótesis de que $\beta_2 = 0$, supusimos tácitamente que la prueba se basaba en una muestra diferente de la utilizada en la prueba de significancia de $\hat{\beta}_3$ conforme a la hipótesis nula de que $\beta_3 = 0$. Pero para probar la hipótesis conjunta de (8.4.1), si empleamos los mismos datos muestrales, violaremos el supuesto del procedimiento de pruebas.[4] El asunto puede plantearse de otra forma:

[3] Para este ejemplo, el valor de asimetría es de 0.2276, y el de curtosis, de 2.9488. Recuerde que para una variable normalmente distribuida los valores de asimetría y curtosis son 0 y 3, respectivamente.

[4] En cualquier muestra dada la cov ($\hat{\beta}_2$, $\hat{\beta}_3$) puede no ser cero; es decir, $\hat{\beta}_2$ y $\hat{\beta}_3$ pueden estar correlacionadas. Véase (7.4.17.)

en (8.3.2) establecimos un intervalo de confianza a 95% para $\beta_2$. Pero si utilizamos los mismos datos muestrales para establecer un intervalo de confianza para $\beta_3$, es decir, con un coeficiente de confianza de 95%, no es posible asegurar que $\beta_2$ y $\beta_3$ se encuentren dentro de sus respectivos intervalos de confianza con una probabilidad de $(1-\alpha)(1-\alpha) = (0.95)(0.95)$.

En otras palabras, aunque las afirmaciones

$$\Pr\,[\hat{\beta}_2 - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2)] = 1-\alpha$$
$$\Pr\,[\hat{\beta}_3 - t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_3) \le \beta_3 \le \hat{\beta}_3 + t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_3)] = 1-\alpha$$

son individualmente ciertas, *no es cierto* que la probabilidad de que $\beta_2$ y $\beta_3$ se encuentren al mismo tiempo en los intervalos

$$[\hat{\beta}_2 \pm t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_2),\; \hat{\beta}_3 \pm t_{\alpha/2}\,\text{ee}\,(\hat{\beta}_3)]$$

sea $(1-\alpha)^2$, porque los intervalos pueden no ser independientes cuando se derivan con la misma información. Para plantear el asunto de otra forma,

> . . . probar una serie de hipótesis simples [individuales] *no* equivale a probar las mismas hipótesis en forma conjunta. La razón intuitiva para esto es que, en una prueba conjunta de varias hipótesis, cualquier hipótesis simple se ve "afectada" por la información de las demás hipótesis.[5]

El resultado final del argumento anterior es que, para un ejemplo dado (muestra), sólo se obtiene un intervalo de confianza o una prueba de significancia. ¿Cómo, entonces, probar la hipótesis nula simultánea de que $\beta_2 = \beta_3 = 0$? En seguida responderemos esta pregunta.

## El método del análisis de varianza en las pruebas de significancia general de una regresión múltiple observada: la prueba *F*

Por las razones recién explicadas, no podemos utilizar la prueba *t* usual para probar la hipótesis conjunta de que los verdaderos coeficientes parciales de pendiente sean simultáneamente iguales a cero. Sin embargo, esta hipótesis conjunta se prueba con la técnica del **análisis de varianza** (ANOVA), presentada en la sección 5.9, lo cual se demuestra de la siguiente manera.

Recuerde la identidad

$$\underbrace{\sum y_i^2}_{\text{SCT}} = \underbrace{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}_{\text{SCE}} + \underbrace{\sum \hat{u}_i^2}_{\text{SCR}} \tag{8.4.2}$$

SCT tiene, como es usual, $n-1$ gl, y SCR tiene $n-3$ gl, por las razones ya expuestas. SCE tiene 2 gl, pues es función de $\hat{\beta}_2$ y $\hat{\beta}_3$. Por consiguiente, según el procedimiento ANOVA estudiado en la sección 5.9, se elabora la tabla 8.1.

Ahora puede demostrarse[6] que, según el supuesto de la distribución normal para $u_i$ y la hipótesis nula $\beta_2 = \beta_3 = 0$, la variable

$$F = \frac{\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)/2}{\sum \hat{u}_i^2/(n-3)} = \frac{\text{SCE/gl}}{\text{SCR/gl}} \tag{8.4.3}$$

se distribuye como la distribución *F* con 2 y $n-3$ gl.

[5] Thomas B. Fomby, R. Carter Hill y Stanley R. Johnson, *Advanced Econometric Methods*, Springer-Verlag, Nueva York, 1984, p. 37.

[6] Véase K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, Nueva York, 1960, pp. 278-280.

**TABLA 8.1**
**Tabla ANOVA para la regresión con tres variables**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| Debido a la regresión (SCE) | $\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$ | 2 | $\dfrac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$ |
| Debido a residuos (SCR) | $\sum \hat{u}_i^2$ | $n-3$ | $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-3}$ |
| Total | $\sum y_i^2$ | $n-1$ | |

¿Qué utilidad puede tener la razón $F$ anterior? Puede demostrarse[7] que, según el supuesto de que los $u_i \sim N(0, \sigma^2)$,

$$E\frac{\sum \hat{u}_i^2}{n-3} = E(\hat{\sigma}^2) = \sigma^2 \tag{8.4.4}$$

Con el supuesto adicional que $\beta_2 = \beta_3 = 0$ se demuestra que

$$\frac{E\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)}{2} = \sigma^2 \tag{8.4.5}$$

Por consiguiente, si la hipótesis nula es verdadera, las ecuaciones (8.4.4) y (8.4.5) proporcionan estimaciones idénticas del verdadero $\sigma^2$. Esta afirmación no debe sorprender, pues, si existe una relación trivial entre $Y$ y $X_2$ y $X_3$, la única fuente de variación en $Y$ se debe a las fuerzas aleatorias representadas por $u_i$. Sin embargo, si la hipótesis nula es falsa, es decir, si $X_2$ y $X_3$ definitivamente ejercen influencia sobre $Y$, no se mantendrá la igualdad entre (8.4.4) y (8.4.5). En este caso, la SCE será relativamente más grande que la SCR, teniendo en cuenta sus respectivos gl. Por consiguiente, el valor $F$ de (8.4.3) proporciona una prueba de la hipótesis nula de que los verdaderos coeficientes de pendiente son simultáneamente cero. Si el valor $F$ calculado de (8.4.3) excede el valor $F$ crítico de la tabla $F$ en el nivel de significancia de $\alpha$%, se rechaza $H_0$; de otra forma no se rechaza. Por otra parte, si el valor $p$ del $F$ observado es lo bastante bajo, podemos rechazar $H_0$.

La tabla 8.2 resume la prueba $F$. De regreso al ejemplo ilustrativo, obtenemos la tabla ANOVA, como se muestra en la tabla 8.3.

**TABLA 8.2**
**Resumen del estadístico *F***

| Hipótesis nula $H_0$ | Hipótesis alternativa $H_1$ | Región crítica. Se rechaza $H_0$ si |
|---|---|---|
| $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 > \sigma_2^2$ | $\dfrac{S_1^2}{S_2^2} > F_{\alpha, n\text{gl}, d\text{gl}}$ |
| $\sigma_1^2 = \sigma_2^2$ | $\sigma_1^2 \neq \sigma_2^2$ | $\dfrac{S_1^2}{S_2^2} > F_{\alpha/2, n\text{gl}, d\text{gl}}$ o $< F_{(1-\alpha/2), n\text{gl}, d\text{gl}}$ |

*Notas:*
1. $\sigma_1^2$ y $\sigma_2^2$ son las dos varianzas poblacionales.
2. $S_1^2$ y $S_2^2$ son las dos varianzas muestrales.
3. $n$ gl y $d$ gl denotan, respectivamente, el numerador y el denominador gl.
4. Al calcular la razón $F$, coloque el valor $S^2$ más grande en el numerador.
5. Los valores críticos $F$ se dan en la última columna. El primer subíndice de $F$ es el nivel de significancia, y los siguientes son los gl del numerador y del denominador.
6. Observe que $F_{(1-\alpha/2), n\text{gl}, d\text{gl}} = 1/F_{\alpha/2, d\text{gl}, n\text{gl}}$.

[7] Véase K. A. Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, Nueva York, 1960, pp. 278-280.

**TABLA 8.3**
**Tabla ANOVA para el ejemplo de mortalidad infantil**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| Debido a la regresión | 257 362.4 | 2 | 128 681.2 |
| Debido a residuos | 106 315.6 | 61 | 1 742.88 |
| Total | 363 678 | 63 | |

Con (8.4.3) obtenemos

$$F = \frac{128\,681.2}{1\,742.88} = 73.8325 \tag{8.4.6}$$

El valor $p$, al obtener un valor $F$ igual o mayor que 73.8325, es casi cero, lo cual implica el rechazo de la hipótesis que establece que el PIBPC y la TAM, conjuntamente, no tienen efecto sobre la mortalidad infantil. Si empleamos el nivel usual de significancia de 5%, el valor $F$ crítico para 2 gl en el numerador y 60 gl en el denominador (sin embargo, los gl reales son 61) es de casi 3.15, o de 4.98 más o menos, si utiliza el nivel de significancia de 1%. Obvio, el valor observado $F$ de casi 74 excede por mucho cualquiera de estos valores críticos $F$.

El procedimiento de prueba $F$ se generaliza de la siguiente manera.

## Prueba de significancia general de una regresión múltiple: la prueba *F*

***Regla de decisión***

Con el modelo de regresión con $k$ variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

Para probar la hipótesis

$$H_0\colon \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

(es decir, todos los coeficientes de pendiente son simultáneamente cero) frente a

$H_1$: no todos los coeficientes de pendiente son simultáneamente cero

calcule

$$F = \frac{\text{SCE/gl}}{\text{SCR/gl}} = \frac{\text{SCE}/(k-1)}{\text{SCR}/(n-k)} \tag{8.4.7}$$

si $F > F_\alpha(k-1, n-k)$, rechace $H_0$; de lo contrario, no la rechace, donde $F_\alpha(k-1, n-k)$ es el valor *F crítico* en el nivel de significancia $\alpha$, y $(k-1)$ gl en el numerador y $(n-k)$ gl en el denominador. Por otra parte, si el valor $p$ del $F$ obtenido de (8.4.7) es lo bastante bajo, se puede rechazar $H_0$.

Sobra decir que en el caso de tres variables ($Y$ y $X_2$, $X_3$) $k$ es 3, en el caso de cuatro variables $k$ es 4, y así sucesivamente.

A propósito, observe que la mayoría de los paquetes de regresión calcula el valor $F$ (dado en la tabla de análisis de varianza) junto con los estadísticos usuales de regresión, como los coeficientes estimados, sus errores estándar, los valores $t$, etc. Suele suponerse que la hipótesis nula para el cálculo $t$ es $\beta_i = 0$.

### *Prueba de hipótesis individual y conjunta*

En la sección 8.3 analizamos la prueba de significancia de un solo coeficiente de regresión, y en la sección 8.4, la prueba de significancia conjunta o general de la regresión estimada (es decir, todos los coeficientes de pendiente son simultáneamente iguales a cero). **Repetimos que estas pruebas son diferentes**. Así, con base en la prueba $t$ o intervalo de confianza (de la sección 8.3), es posible aceptar la hipótesis de que un coeficiente de pendiente particular, $\beta_k$, es cero, y aun rechazar la hipótesis conjunta de que todos los coeficientes de pendiente son cero.

> La lección es que el "mensaje" conjunto de los intervalos de confianza individuales no sustituye una región de confianza conjunta [implicada por la prueba $F$] en el momento de realizar pruebas de hipótesis conjuntas y efectuar afirmaciones de confianza conjuntas.[8]

## Una relación importante entre $R^2$ y $F$

Existe una relación estrecha entre el coeficiente de determinación $R^2$ y la prueba $F$ en el análisis de varianza. En el supuesto de que las perturbaciones $u_i$ están normalmente distribuidas, y según la hipótesis nula de que $\beta_2 = \beta_3 = 0$, vimos que

$$F = \frac{\text{SCE}/2}{\text{SCR}/(n-3)} \tag{8.4.8}$$

sigue una distribución $F$ con 2 y $n - 3$ gl.

De modo más general, en el caso de $k$ variables (inclusive el intercepto), si suponemos que las perturbaciones están normalmente distribuidas y que la hipótesis nula es

$$H_0\colon \beta_2 = \beta_3 = \cdots = \beta_k = 0 \tag{8.4.9}$$

entonces se cumple que

$$F = \frac{\text{SCE}/(k-1)}{\text{SCR}/(n-k)} \qquad \textbf{(8.4.7) = (8.4.10)}$$

sigue la distribución $F$ con $k - 1$ y $n - k$ gl. (*Nota:* El número total de parámetros que se va a estimar es $k$, de los cuales uno es el término de intercepto.)

Manipulamos (8.4.10) de la siguiente manera:

$$\begin{aligned} F &= \frac{n-k}{k-1}\frac{\text{SCE}}{\text{SCR}} \\ &= \frac{n-k}{k-1}\frac{\text{SCE}}{\text{SCT}-\text{SCE}} \\ &= \frac{n-k}{k-1}\frac{\text{SCE}/\text{SCT}}{1-(\text{SCE}/\text{SCT})} \\ &= \frac{n-k}{k-1}\frac{R^2}{1-R^2} \\ &= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \end{aligned} \tag{8.4.11}$$

[8] Fomby *et al., op. cit.*, p. 42.

**TABLA 8.4**
**Tabla ANOVA en términos de $R^2$**

| Origen de la variación | SC | gl | SCM* |
|---|---|---|---|
| Debido a la regresión | $R^2(\sum y_i^2)$ | 2 | $R^2(\sum y_i^2)/2$ |
| Debido a residuos | $(1 - R^2)(\sum y_i^2)$ | $n - 3$ | $(1 - R^2)(\sum y_i^2)/(n - 3)$ |
| Total | $\sum y_i^2$ | $n - 1$ | |

*Observe que para calcular el valor $F$ no hay necesidad de multiplicar $R^2$ y $(1 - R^2)$ por $\sum y_i^2$ porque este término desaparece, como se muestra en la ecuación (8.4.12).

donde se emplea la definición $R^2 = \text{SCE/SCT}$. La ecuación (8.4.11) muestra la forma en que $F$ y $R^2$ están relacionadas. Las dos varían en relación directa. Cuando $R^2 = 0$, $F$ es cero *ipso facto*. Cuanto mayor sea la $R^2$, mayor será el valor $F$. En el límite, cuando $R^2 = 1$, $F$ es infinita. *Así, la prueba F, que mide la significancia general de la regresión estimada, es también una prueba de significancia de $R^2$.* En otras palabras, la prueba de la hipótesis nula (8.4.9) equivale a probar la hipótesis nula de que el $R^2$ (poblacional) es cero.

Para el caso de tres variables (8.4.11) se convierte en

$$F = \frac{R^2/2}{(1 - R^2)/(n - 3)} \tag{8.4.12}$$

En virtud de la estrecha conexión entre $F$ y $R^2$, la tabla ANOVA (tabla 8.1) se reformula como la tabla 8.4.

Para nuestro ejemplo ilustrativo, al utilizar (8.4.12) obtenemos:

$$F = \frac{0.7077/2}{(1 - 0.7077)/61} = 73.8726$$

que es casi el mismo que obtuvimos antes, excepto por los errores de redondeo.

Una ventaja de la prueba $F$ expresada en términos de $R^2$ es su facilidad de cálculo: todo lo que se necesita saber es el valor de $R^2$. Por consiguiente, la prueba de significancia general $F$ dada en (8.4.7) se expresa en términos de $R^2$ como se indica en la tabla 8.4.

## Prueba de significancia general de una regresión múltiple en términos de $R^2$

***Regla de decisión***

Para probar la significancia general de una regresión en términos de $R^2$: una prueba alterna pero equivalente a (8.4.7).

Dado el modelo de regresión con $k$ variables:

$$Y_i = \beta_i + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_x X_{ki} + u_i$$

Para probar la hipótesis

$$H_0\text{: } \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

frente a

$H_1$: no todos los coeficientes de pendiente son simultáneamente cero

calcular

$$F = \frac{R^2/(k - 1)}{(1 - R^2)/(n - k)} \tag{8.4.13}$$

Si $F > F_{\alpha(k-1,\, n-k)}$, rechace $H_0$; de lo contrario, puede aceptar $H_0$, donde $F_{\alpha(k-1,\, n-k)}$ es el valor $F$ crítico en el nivel de significancia $\alpha$, y $(k - 1)$ gl es el numerador y $(n - k)$ gl el denominador. Por otra parte, si el valor $p$ del $F$ obtenido de (8.4.13) es lo bastante bajo, rechace $H_0$.

Antes de continuar, reconsideremos el ejemplo 7.5 del capítulo 7. De la regresión (7.10.7) se observa que el PIBPC (relativo al PIB *per cápita*) y el PIBPC al cuadrado explican sólo 10.92% de la variación en la tasa de crecimiento del PIB para una muestra de 190 países. Esta $R^2$ de 0.1092 parece un valor "bajo". ¿Es estadísticamente distinto de cero? ¿Cómo averiguarlo?

Recuerde el análisis anterior en el apartado "Una relación importante entre $R^2$ y $F$", respecto de la relación entre $R^2$ y el valor $F$ como se dio en (8.4.11) u (8.4.12) para el caso específico de dos regresoras. Como se observó, si $R^2$ es cero, $F$ es cero *ipso facto*, lo cual será el caso si las regresoras no tienen ningún efecto sobre la regresada. Por tanto, si añadimos $R^2 = 0.1092$ en la fórmula (8.4.12), obtenemos

$$F = \frac{0.1092/2}{(1 - 0.1092)/187} = 11.4618 \tag{8.4.13}$$

Según la hipótesis nula de que $R^2 = 0$, el valor $F$ anterior sigue la distribución $F$ con 2 y 187 gl en el numerador y denominador, respectivamente. (*Nota:* Hay 190 observaciones y dos regresoras.) De la tabla $F$ se observa que este valor $F$ es significativo en un nivel de casi 5%; de hecho, el valor $p$ es 0.00002. Por tanto, podemos rechazar la hipótesis nula de que las dos regresoras no tienen efecto sobre la regresada, a pesar de que $R^2$ es de sólo 0.1092.

Este ejemplo permite una importante observación empírica: en los datos transversales que implican varias observaciones, por lo general se obtienen $R^2$ bajas, debido a la diversidad de las unidades transversales. En consecuencia, no hay que sorprenderse o preocuparse si se tiene una $R^2$ baja en regresiones transversales. Lo relevante es que el modelo esté bien especificado, que las regresoras tengan los signos correctos (es decir, los teóricamente esperados) y que (con un poco de suerte) los coeficientes de regresión sean estadísticamente significativos. El lector debe verificar que, *de manera individual*, ambas regresoras de (7.10.7) sean estadísticamente significativas en un nivel de 5% o mejor (es decir, menor que 5%).

## La contribución "incremental" o "marginal" de una variable explicativa

En el capítulo 7 afirmamos que por lo general no podemos asignar el valor $R^2$ entre las distintas regresoras. En el ejemplo de la mortalidad infantil descubrimos que la $R^2$ valía 0.7077, pero no fue posible saber qué parte de dicho valor se debía a la regresora PIBPC y qué parte a la TAM, debido a la posible correlación entre las dos regresoras de las muestras disponibles. Se aclara este asunto con la técnica del análisis de varianza.

Para el ejemplo presente, vimos que *individualmente* $X_2$ (PIBPC) y $X_3$ (TAM) eran estadísticamente significativas con base en pruebas $t$ (*separadas*). También observamos que, con base en la prueba $F$, *colectivamente* las dos regresoras tienen un efecto importante sobre la regresada $Y$ (mortalidad infantil).

Ahora suponga que introducimos PIBPC y TAM *de manera secuencial*; es decir, primero hacemos la regresión de la mortalidad infantil sobre PIBPC y evaluamos su significancia, y después añadimos TAM al modelo para averiguar si contribuye en algo (por supuesto, puede invertirse el orden en el que se añaden PIBPC y TAM). Por contribución nos referimos al posible incremento "significativo" de la SCE (y por tanto de la $R^2$) en relación con la SCR si añadimos una variable al modelo. Esta contribución puede llamarse con toda propiedad contribución **incremental** o **marginal** de una variable explicativa.

El tema de la contribución incremental es importante en la práctica. En la mayoría de las investigaciones empíricas, el investigador puede no estar por completo seguro de que se justifique agregar una variable $X$ al modelo, pues ya hay en él muchas otras variables $X$. No se desea incluir variable(s) cuya contribución a la SCE sea poca. Por la misma causa, tampoco se quiere excluir variable(s) que aumente($n$) sustancialmente la SCE. Pero, ¿cómo decidir si una variable $X$ reduce significativamente la SCR? La técnica del análisis de varianza puede extenderse con facilidad para responder esta pregunta.

**TABLA 8.5**
**Tabla ANOVA para la ecuación de regresión (8.4.14)**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| SCE (debido al PIBPC) | 60 449.5 | 1 | 60 449.5 |
| SCR | 303 228.5 | 62 | 4 890.7822 |
| Total | 363 678 | 63 | |

Suponga que primero hace la regresión de la mortalidad infantil sobre el PIBPC y obtiene la siguiente regresión:

$$\begin{aligned} \widehat{\text{MI}}_i &= 157.4244 \quad - \quad 0.0114\ \text{PIBPC} \\ t &= (15.9894) \quad (-3.5156) \qquad r^2 = 0.1662 \\ \text{valor } p &= \quad (0.0000) \qquad (0.0008) \qquad r^2 \text{ ajust.} = 0.1528 \end{aligned} \tag{8.4.14}$$

Como muestran estos resultados, PIBPC tiene un efecto importante sobre MI. La tabla ANOVA correspondiente a la regresión anterior se da en la tabla 8.5.

Si suponemos que las perturbaciones $u_i$ están normalmente distribuidas y establecemos la hipótesis de que PIBPC no tiene ningún efecto sobre MI, obtenemos el valor $F$ de

$$F = \frac{60\,449.5}{4\,890.7822} = 12.3598 \tag{8.4.15}$$

el cual sigue la distribución $F$ con 1 y 62 gl. Este valor $F$ resulta muy significativo, pues el valor $p$ calculado es de 0.0008. Por tanto, como antes, rechazamos la hipótesis de que PIBPC no tiene efecto alguno sobre MI. A propósito, observe que $t^2 = (-3.5156)^2 = 12.3594$, lo cual es casi lo mismo que el valor $F$ de (8.4.15), donde el valor $t$ se obtiene de (8.4.14). Pero esto no debe sorprender, porque el cuadrado del estadístico $t$ con $n$ gl es igual al valor $F$ con 1 gl en el numerador y $n$ gl en el denominador, relación que se estableció en el capítulo 5. Note que en este ejemplo $n = 64$.

Una vez efectuada la regresión (8.4.14), supongamos que decidimos agregar $X_3$ al modelo y obtener la regresión múltiple (8.1.4). Las preguntas son:

1. ¿Cuál es la contribución marginal o incremental de TAM si sabemos que PIBPC ya aparece en el modelo y está relacionada significativamente con MI?
2. ¿Es la contribución incremental de TAM estadísticamente significativa?
3. ¿Cuál es el criterio para agregar variables al modelo?

Estas preguntas se resuelven mediante la técnica ANOVA. Para ver esto elaboramos la tabla 8.6. En esta tabla, $X_2$ se refiere al PIBPC y $X_3$ a la TAM.

Para evaluar la contribución *incremental* de $X_3$ después de considerar la contribución de $X_2$, calculamos

$$\begin{aligned} F &= \frac{Q_2/\text{gl}}{Q_4/\text{gl}} \\ &= \frac{(\text{SCE}_{\text{nueva}} - \text{SCE}_{\text{vieja}})/\text{número de regresoras nuevas}}{\text{SCR}_{\text{nueva}}/\text{gl}\ (= n - \text{número de parámetros en el nuevo modelo})} \\ &= \frac{Q_2/1}{Q_4/61} \text{ para este ejemplo} \end{aligned} \tag{8.4.16}$$

**TABLA 8.6**
**Tabla ANOVA para evaluar la contribución incremental de una o más variables**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| SCE debido sólo a $X_2$ | $Q_1 = \hat{\beta}_{12}^2 \sum x_2^2$ | 1 | $\frac{Q_1}{1}$ |
| SCE debido a la adición de $X_3$ | $Q_2 = Q_3 - Q_1$ | 1 | $\frac{Q_2}{1}$ |
| SCE debido tanto a $X_2$ como a $X_3$ | $Q_3 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$ | 2 | $\frac{Q_3}{2}$ |
| SCR | $Q_4 = Q_5 - Q_3$ | $n-3$ | $\frac{Q_4}{n-3}$ |
| Total | $Q_5 = \sum y_i^2$ | $n-1$ | |

donde $SCE_{nueva}$ = SCE según el modelo nuevo (es decir, después de agregar las regresoras nuevas = $Q_3$), $SCE_{vieja}$ = SCE según el modelo antiguo ( = $Q_1$) y $SCR_{nueva}$ = SCR según el nuevo (es decir, después de considerar todas las regresoras = $Q_4$). Para este ejemplo ilustrativo, los resultados se presentan en la tabla 8.7.

Ahora, al aplicar la ecuación (8.4.16), obtenemos

$$F = \frac{196\,912.9}{1\,742.8786} = 112.9814 \tag{8.4.17}$$

Según los supuestos usuales, este valor $F$ sigue la distribución $F$ con 1 y 62 gl. El lector verificará que este valor $F$ es muy significativo, lo cual sugiere que la incorporación de TAM al modelo aumenta en gran medida la SCE y, por tanto, el valor $R^2$. En consecuencia, TAM debe incorporarse al modelo. Una vez más, observe que si elevamos al cuadrado el valor del coeficiente de TAM en la regresión múltiple (8.1.4), que es $(-10.6293)^2$, obtendremos el valor $F$ de (8.4.17), salvo por errores de redondeo.

A propósito, la razón $F$ de (8.4.16) se reformula sólo en términos de los valores $R^2$, como en (8.4.13). Como lo indica el ejercicio 8.2, la razón $F$ de (8.4.16) es *equivalente* a la siguiente razón $F$:[9]

$$\begin{aligned} F &= \frac{\left(R^2_{\text{nueva}} - R^2_{\text{vieja}}\right)/\text{gl}}{\left(1 - R^2_{\text{nueva}}\right)/\text{gl}} \\ &= \frac{\left(R^2_{\text{nueva}} - R^2_{\text{vieja}}\right)/\text{número de regresoras nuevas}}{\left(1 - R^2_{\text{nueva}}\right)/\text{gl}\ (= n - \text{número de parámetros en el nuevo modelo})} \end{aligned} \tag{8.4.18}$$

**TABLA 8.7**
**Tabla ANOVA para el ejemplo ilustrativo: análisis incremental**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| SCE debido a PIBPC | 60 449.5 | 1 | 60 449.5 |
| SCE debido a la adición de TAM | 196 912.9 | 1 | 196 912.9 |
| ESCE debido tanto a PIBPC como a TAM | 257 362.4 | 2 | 128 681.2 |
| SCR | 106 315.6 | 63 | 1 742.8786 |
| Total | 363 678 | 63 | |

[9] La siguiente prueba $F$ es un caso especial de la prueba $F$ más general dada en (8.6.9) u (8.6.10), en la sección 8.6.

Esta razón $F$ también sigue la distribución $F$ con los gl apropiados en el numerador y el denominador, 1 y 61, respectivamente, en el ejemplo ilustrativo.

Para nuestro ejemplo, $R^2_{\text{nueva}} = 0.7077$ [de (8.1.4)] y $R^2_{\text{vieja}} = 0.1662$ [de (8.4.14)]. Por consiguiente,

$$F = \frac{(0.7077 - 0.1662)/1}{(1 - 0.7077)/61} = 113.05 \tag{8.4.19}$$

lo cual es casi lo mismo que el valor $F$ de (8.4.17), excepto por errores de redondeo. Esta $F$ es muy significativa, pues refuerza el descubrimiento anterior de que la variable TAM pertenece al modelo.

*Advertencia:* Si utiliza la versión $R^2$ de la prueba $F$ dada en (8.4.11), asegúrese de que la variable dependiente en el modelo reciente y en el anterior sea la misma. Si son diferentes utilice la prueba $F$ de (8.4.16).

### *Cuándo agregar una nueva variable*

El procedimiento de prueba $F$ recién presentado constituye un método formal para decidir si una variable debe adicionarse a un modelo de regresión. A menudo, los investigadores se enfrentan a la labor de escoger entre diversos modelos en competencia, **que consideran la misma variable dependiente** pero diferentes variables explicativas. Como un asunto de elección *ad hoc* (debido a que muchas veces la base teórica del análisis es débil), estos investigadores seleccionan el modelo que presenta la $R^2$ ajustada más elevada. Por consiguiente, si la inclusión de una variable incrementa $\bar{R}^2$, ésta se conserva en el modelo aunque no reduzca significativamente la SCR en sentido estadístico. La pregunta entonces es: ¿cuándo aumenta $R^2$ ajustada? Puede demostrarse que *$\bar{R}^2$ aumenta si el valor $t$ del coeficiente de la variable recién agregada es mayor que 1 en valor absoluto*, donde el valor $t$ se calcula según la hipótesis de que el valor poblacional del mencionado coeficiente es cero [es decir, el valor $t$ calculado a partir de (5.3.2) conforme a la hipótesis de que el verdadero $\beta$ es cero].[10] El criterio anterior también se plantea en forma diferente: *$\bar{R}^2$ aumenta con la adición de una variable explicativa sólo si el valor $F(= t^2)$ de esa variable es superior a 1.*

Con cualquier criterio, la variable TAM con un valor $t$ de $-10.6293$ o un valor $F$ de 112.9814 debe aumentar $\bar{R}^2$, lo cual sin duda sucede: cuando se agrega TAM al modelo, $\bar{R}^2$ se incrementa de 0.1528 a 0.6981.

### *Cuándo agregar un grupo de variables*

¿Se puede elaborar una regla similar para decidir si se justifica agregar (o quitar) un grupo de variables del modelo? La respuesta debe ser clara a partir de (8.4.18): *si la adición (o eliminación) de un grupo de variables al modelo genera un valor F mayor (menor) que 1, $R^2$ aumenta (disminuye)*. Por supuesto, de (8.4.18) se deriva fácilmente si la adición (disminución) de un grupo de variables incrementa (reduce) de modo significativo el poder explicativo de un modelo de regresión.

## 8.5 Prueba de igualdad de dos coeficientes de regresión

Suponga que en la regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \tag{8.5.1}$$

[10] Para una demostración, véase Dennis J. Aigner, *Basic Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1971, pp. 91-92.

deseamos probar la hipótesis

$$H_0: \beta_3 = \beta_4 \quad \text{o} \quad (\beta_3 - \beta_4) = 0$$
$$H_1: \beta_3 \neq \beta_4 \quad \text{o} \quad (\beta_3 - \beta_4) \neq 0 \tag{8.5.2}$$

es decir, los dos coeficientes de pendiente $\beta_3$ y $\beta_4$ son iguales.

Una hipótesis nula de este tipo tiene importancia práctica; por ejemplo, si (8.5.1) representa la función de demanda de un bien donde $Y$ = cantidad demandada del bien, $X_2$ = precio del bien, $X_3$ = ingreso del consumidor y $X_4$ = riqueza del consumidor. La hipótesis nula en este caso significa que los coeficientes de ingreso y riqueza son los mismos. Si $Y_i$ y las $X$ están expresadas en forma logarítmica, la hipótesis nula en (8.5.2) implica que las elasticidades ingreso y riqueza del consumo son iguales. (¿Por qué?)

¿Cómo probamos una hipótesis nula como ésta? Con los supuestos clásicos se demuestra que

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{ee}\,(\hat{\beta}_3 - \hat{\beta}_4)} \tag{8.5.3}$$

sigue la distribución $t$ con $(n - 4)$ gl porque (8.5.1) es un modelo con cuatro variables o, más generalmente, con $(n - k)$ gl, donde $k$ es el número total de parámetros estimados, inclusive el término constante. El ee $(\hat{\beta}_3 - \hat{\beta}_4)$ se obtiene de la siguiente fórmula conocida (véase detalles en el **apéndice A**):

$$\text{ee}\,(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\text{var}\,(\hat{\beta}_3) + \text{var}\,(\hat{\beta}_4) - 2\,\text{cov}\,(\hat{\beta}_3, \hat{\beta}_4)} \tag{8.5.4}$$

Si sustituimos la hipótesis nula y la expresión para el ee $\hat{\beta}_3 - \hat{\beta}_4$ en (8.5.3), el estadístico de prueba se convierte en

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}\,(\hat{\beta}_3) + \text{var}\,(\hat{\beta}_4) - 2\,\text{cov}\,(\hat{\beta}_3, \hat{\beta}_4)}} \tag{8.5.5}$$

Ahora el procedimiento de prueba comprende los siguientes pasos:

1. Estime $\hat{\beta}_3$ y $\hat{\beta}_4$. Cualquier software estándar puede hacerlo.
2. La mayoría de los programas de computación calcula las varianzas y covarianzas de los parámetros estimados.[11] De estas estimaciones, el error estándar en el denominador de (8.5.5) se obtiene fácilmente.
3. Obtenga la razón $t$ de (8.5.5). Observe que la hipótesis nula en el caso presente es $(\beta_3 - \beta_4) = 0$.
4. Si la variable $t$ calculada de (8.5.5) excede el valor $t$ crítico en el nivel de significancia designado para los gl dados, puede rechazar la hipótesis nula; de lo contrario, no la rechace. En forma alterna, si el valor $p$ del estadístico $t$ de (8.5.5) es razonablemente bajo, puede rechazar la hipótesis nula. Observe que mientras más bajo sea el valor $p$, mayor será la evidencia en contra de la hipótesis nula. Por tanto, cuando se dice que un valor $p$ es bajo o razonablemente bajo, se quiere decir que es menor que el nivel de significancia, por ejemplo, 10, 5 o 1%. En esta decisión está implícito cierto criterio personal.

[11] La expresión algebraica para la fórmula de la covarianza es un poco compleja. En el **apéndice C** se ofrece una expresión compacta de ella, aunque con notación matricial.

**EJEMPLO 8.2**
*De nuevo, la función cúbica de costo*

Recuerde la función cúbica del costo total estimada en el ejemplo 7.4, sección 7.10, que se reproduce en seguida:

$$\begin{aligned} \hat{Y}_i &= 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ \text{ee} &= \quad (6.3753) \qquad (4.7786) \qquad (0.9857) \qquad (0.0591) \\ & \text{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.0576; \qquad R^2 = 0.9983 \end{aligned} \tag{7.10.6}$$

donde $Y$ es el costo total y $X$ es la producción, y donde las cifras en paréntesis son los errores estándar estimados.

Suponga que deseamos probar la hipótesis de que los coeficientes de los términos $X^2$ y $X^3$ en la función cúbica de costo son los mismos, es decir, $\beta_3 = \beta_4$ o $(\beta_3 - \beta_4) = 0$. En la regresión (7.10.6) aparecen todos los resultados necesarios para realizar la prueba $t$ a partir de (8.5.5). La mecánica es la siguiente:

$$\begin{aligned} t &= \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\,\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \\ &= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} \\ &= \frac{-13.9011}{1.0442} = -13.3130 \end{aligned} \tag{8.5.6}$$

El lector puede verificar que, para 6 gl (¿por qué?), el valor $t$ observado excede el valor $t$ crítico aun en el nivel de significancia de 0.002 (o 0.2%) (prueba de dos colas); el valor $p$ es extremadamente pequeño, 0.000006. Por tanto, podemos rechazar la hipótesis de que los coeficientes de $X^2$ y $X^3$ en la función cúbica de costo son idénticos.

## 8.6 Mínimos cuadrados restringidos: pruebas de restricciones de igualdades lineales

Hay ocasiones en las cuales la teoría económica puede sugerir que los coeficientes en un modelo de regresión satisfacen algunas restricciones de igualdades lineales. Por ejemplo, considere la función de producción Cobb-Douglas:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{7.9.1) = (8.6.1}$$

donde $Y$ = producción, $X_2$ = insumo trabajo y $X_3$ = insumo capital. Escrita en forma de logaritmo, la ecuación se convierte en

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \tag{8.6.2}$$

donde $\beta_0 = \ln \beta_1$.

Ahora, si existen rendimientos constantes a escala (un cambio equiproporcional en la producción ante un cambio equiproporcional en los insumos), la teoría económica sugeriría que

$$\beta_2 + \beta_3 = 1 \tag{8.6.3}$$

el cual es un ejemplo de una restricción de igualdad lineal.[12]

¿Cómo saber si hay rendimientos constantes a escala, es decir, si es válida la restricción (8.6.3)? Existen dos métodos.

[12] Si tuviéramos $\beta_2 + \beta_3 < 1$, esta relación sería un ejemplo de restricción de desigualdad lineal. Para el manejo de tales restricciones se requieren técnicas de programación matemática.

## El enfoque de la prueba *t*

El procedimiento más simple es estimar (8.6.2) de la forma usual sin tener en cuenta la restricción (8.6.3) explícitamente. Esto se denomina **regresión no restringida** o **no limitada**. Tras estimar $\beta_2$ y $\beta_3$ (por ejemplo, mediante el método de MCO) se realiza una prueba de la hipótesis o restricción (8.6.3) mediante la prueba *t* de (8.5.3), a saber:

$$\begin{aligned} t &= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{\text{ee}\,(\hat{\beta}_2 + \hat{\beta}_3)} \\ &= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\text{var}\,(\hat{\beta}_2) + \text{var}\,(\hat{\beta}_3) + 2\,\text{cov}\,(\hat{\beta}_2, \hat{\beta}_3)}} \end{aligned} \tag{8.6.4}$$

donde $(\beta_2 + \beta_3) = 1$ según la hipótesis nula y el denominador es el error estándar de $(\hat{\beta}_2 + \hat{\beta}_3)$. Entonces, según la sección 8.5, si el valor *t* calculado a partir de (8.6.4) excede el valor *t* crítico en el nivel de significancia seleccionado, rechazamos la hipótesis de rendimientos constantes a escala; de lo contrario, no la rechazamos.

## Enfoque de la prueba *F*: mínimos cuadrados restringidos

La prueba *t* anterior es una clase de examen *post mortem*, pues se trata de encontrar si se satisface la restricción lineal después de estimar la regresión "no restringida". Un método directo sería incorporar la restricción (8.6.3) en el procedimiento de estimación desde el comienzo. En el ejemplo presente, este procedimiento se realiza fácilmente. De (8.6.3) vemos que:

$$\beta_2 = 1 - \beta_3 \tag{8.6.5}$$

o

$$\beta_3 = 1 - \beta_2 \tag{8.6.6}$$

Por consiguiente, con cualquiera de estas igualdades podemos eliminar un coeficiente $\beta$ en (8.6.2) y estimar la ecuación resultante. Así, si utilizamos (8.6.5) escribimos la función de producción Cobb-Douglas de la siguiente manera:

$$\begin{aligned} \ln Y_i &= \beta_0 + (1 - \beta_3) \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \ln X_{2i} + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i \end{aligned}$$

o

$$(\ln Y_i - \ln X_{2i}) = \beta_0 + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i \tag{8.6.7}$$

o

$$\ln (Y_i/X_{2i}) = \beta_0 + \beta_3 \ln (X_{3i}/X_{2i}) + u_i \tag{8.6.8}$$

donde $(Y_i/X_{2i})$ = razón producción/trabajo y $(X_{3i}/X_{2i})$ = razón capital/trabajo, indicadores de gran importancia económica.

Observe cómo se transforma la ecuación original (8.6.2). Una vez que se estima $\beta_3$ de (8.6.7) u (8.6.8), $\beta_2$ se estima con facilidad a partir de la relación (8.6.5). Sobra decir que este procedimiento garantiza que la suma de los coeficientes estimados de los dos insumos iguale a 1. El procedimiento esquematizado en (8.6.7) o en (8.6.8) se conoce como **mínimos cuadrados restringidos (MCR)**. Este procedimiento puede generalizarse a modelos con cualquier número de variables explicativas y más de una restricción de igualdad lineal. La generalización se encuentra en Theil.[13] (Véase en seguida la prueba *F* general.)

[13] Henri Theil, *Principles of Econometrics*, John Wiley & Sons, Nueva York, 1971, pp. 43-45.

¿Cómo comparamos las regresiones de mínimos cuadrados no restringidas y las restringidas? En otras palabras, ¿cómo saber que, por ejemplo, la restricción (8.6.3) es válida? Esta pregunta se verifica al aplicar la prueba $F$ de la siguiente manera. Sea

$\sum \hat{u}^2_{\text{NR}} =$ SCR de la regresión no restringida (8.6.2)

$\sum \hat{u}^2_{\text{R}} =$ SCR de la regresión restringida (8.6.7)

$m =$ número de restricciones lineales (1 en el presente ejemplo)

$k =$ número de parámetros en la regresión no restringida

$n =$ número de observaciones

Entonces,

$$\begin{aligned} F &= \frac{(\text{SCR}_{\text{R}} - \text{SCR}_{\text{NR}})/m}{\text{SCR}_{\text{NR}})(n-k)} \\ &= \frac{\left(\sum \hat{u}^2_{\text{R}} - \sum \hat{u}^2_{\text{NR}}\right)/m}{\sum \hat{u}^2_{\text{NR}}/(n-k)} \end{aligned} \tag{8.6.9}$$

sigue la distribución $F$ con $m$, $(n-k)$ gl. (*Nota:* Las letras NR y R representan "no restringida" y "restringida", respectivamente.)

La anterior prueba $F$ también se expresa en términos de $R^2$ de la siguiente manera:

$$F = \frac{\left(R^2_{\text{NR}} - R^2_{\text{R}}\right)/m}{\left(1 - R^2_{\text{NR}}\right)/(n-k)} \tag{8.6.10}$$

donde $R^2_{\text{NR}}$ y $R^2_{\text{R}}$ son los valores $R^2$ obtenidos de las regresiones no restringida y restringida respectivamente, es decir, de las regresiones (8.6.2) y (8.6.7). Debe observarse que

$$R^2_{\text{NR}} \geq R^2_{\text{R}} \tag{8.6.11}$$

y

$$\sum \hat{u}^2_{\text{NR}} \leq \sum \hat{u}^2_{\text{R}} \tag{8.6.12}$$

En el ejercicio 8.4 se le pide justificar estas afirmaciones.

*Advertencia:* Al utilizar (8.6.10) tenga en mente que si la variable dependiente en los modelos restringido y no restringido no es la misma, $R^2_{\text{NR}}$ y $R^2_{\text{R}}$ no son directamente comparables. En ese caso, utilice el procedimiento descrito en el capítulo 7 para hacer comparables los dos valores de $R^2$ (véase el ejemplo 8.3 en seguida), o utilice la prueba $F$ de la ecuación (8.6.9).

**EJEMPLO 8.3**
*Función de producción Cobb-Douglas para la economía mexicana, 1955-1974*

A fin de ilustrar el análisis anterior, considere los datos de la tabla 8.8. El ajuste de la función de producción Cobb-Douglas a esos datos produjo los siguientes resultados:

$$\begin{aligned} \widehat{\ln \text{PIB}}_t &= -1.6524 + 0.3397 \ln \text{Trabajo}_t + 0.8460 \ln \text{Capital}_t \\ t &= (-2.7259) \quad (1.8295) \quad (9.0625) \\ \text{valor } p &= (0.0144) \quad (0.0849) \quad (0.0000) \\ &R^2 = 0.9951 \qquad \text{SCR}_{\text{NR}} = 0.0136 \end{aligned} \tag{8.6.13}$$

donde $\text{SCR}_{\text{NR}}$ es la SCR no restringida, pues no se pusieron restricciones al estimar (8.6.13).

**TABLA 8.8**
**PIB real, empleo y capital fijo real en México**

Fuente: Victor J. Elias, *Sources of Growth: A Study of Seven Latin American Economies*, International Center for Economic Growth, ICS Press, San Francisco, 1992. Datos tomados de las tablas E5, E12 y E14.

| Año | PIB* | Empleo† | Capital fijo‡ |
|---|---|---|---|
| 1955 | 114 043 | 8 310 | 182 113 |
| 1956 | 120 410 | 8 529 | 193 749 |
| 1957 | 129 187 | 8 738 | 205 192 |
| 1958 | 134 705 | 8 952 | 215 130 |
| 1959 | 139 960 | 9 171 | 225 021 |
| 1960 | 150 511 | 9 569 | 237 026 |
| 1961 | 157 897 | 9 527 | 248 897 |
| 1962 | 165 286 | 9 662 | 260 661 |
| 1963 | 178 491 | 10 334 | 275 466 |
| 1964 | 199 457 | 10 981 | 295 378 |
| 1965 | 212 323 | 11 746 | 315 715 |
| 1966 | 226 977 | 11 521 | 337 642 |
| 1967 | 241 194 | 11 540 | 363 599 |
| 1968 | 260 881 | 12 066 | 391 847 |
| 1969 | 277 498 | 12 297 | 422 382 |
| 1970 | 296 530 | 12 955 | 455 049 |
| 1971 | 306 712 | 13 338 | 484 677 |
| 1972 | 329 030 | 13 738 | 520 553 |
| 1973 | 354 057 | 15 924 | 561 531 |
| 1974 | 374 977 | 14 154 | 609 825 |

* Millones de pesos de 1960.
† Miles de personas.
‡ Millones de pesos de 1960.

Ya vimos en el capítulo 7 cómo interpretar los coeficientes de la función de producción Cobb-Douglas. Como se aprecia, la elasticidad producción/trabajo es de casi 0.34 y la elasticidad producción/capital vale casi 0.85. Si sumamos estos coeficientes se obtiene 1.19, lo que sugiere que quizá la economía mexicana experimentó durante ese periodo establecido rendimientos crecientes a escala. Por supuesto, desconocemos si 1.19 es estadísticamente diferente de 1.

Para ver si es el caso, se impone una restricción de rendimientos constantes a escala, lo cual da la siguiente regresión:

$$\widehat{\ln\,(\text{PIB/Trabajo})}_t = -0.4947 + 1.0153 \ln\,(\text{Capital/Trabajo})_t \qquad \textbf{(8.6.14)}$$

$$t = (-4.0612) \quad (28.1056)$$

$$\text{valor } p = \quad (0.0007) \quad (0.0000)$$

$$R^2_{\text{R}} = 0.9777 \qquad \text{SCR}_{\text{R}} = 0.0166$$

donde $\text{SCR}_{\text{R}}$ es la SCR restringida, pues impusimos la restricción de que haya rendimientos constantes a escala.

(*continúa*)

**EJEMPLO 8.3**
(*continuación*)

Como la variable dependiente en las dos regresiones anteriores es diferente, tenemos que utilizar la prueba $F$ dada en (8.6.9). Se cuenta con los datos necesarios para obtener el valor $F$.

$$F = \frac{(SCR_R - SCR_{NR})/m}{SCR_{NR}/(n-k)}$$

$$= \frac{(0.0166 - 0.0136)/1}{(0.0136)/(20-3)}$$

$$= 3.75$$

Observe que en el presente caso $m = 1$, pues sólo se impuso una restricción y $(n - k)$ es 17, en vista de que se tienen 20 observaciones y tres parámetros en la regresión no restringida.

Este valor $F$ sigue una distribución $F$ con 1 gl en el numerador y 17 en el denominador. El lector puede verificar con facilidad que esta $F$ no es significativa, en un nivel de significancia de 5%. (Véase el **apéndice D**, tabla D.3.)

Así, la conclusión es que la economía mexicana quizá se caracterizó por rendimientos constantes a escala en el periodo de muestra y, por tanto, no hay daño alguno al utilizar la regresión restringida dada en (8.6.14). Como muestra esta regresión, si la razón capital/trabajo se incrementó 1%, en promedio, la productividad del trabajo aumentó casi 1%.

## Prueba *F* general[14]

La prueba $F$ dada en (8.6.10) o su equivalente en (8.6.9) proporciona un método general de prueba de hipótesis sobre uno o más parámetros del modelo de regresión con $k$ variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \qquad \textbf{(8.6.15)}$$

La prueba $F$ de (8.4.16) o la prueba $t$ de (8.5.3) no es otra cosa que una aplicación específica de (8.6.10). Así, hipótesis como

$$H_0\colon \beta_2 = \beta_3 \qquad \textbf{(8.6.16)}$$

$$H_0\colon \beta_3 + \beta_4 + \beta_5 = 3 \qquad \textbf{(8.6.17)}$$

que consideran algunas restricciones lineales en los parámetros del modelo con $k$ variables, o hipótesis como

$$H_0\colon \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \qquad \textbf{(8.6.18)}$$

que implica que algunas regresoras están ausentes del modelo, pueden probarse mediante la prueba $F$ de (8.6.10).

Del análisis en las secciones 8.4 y 8.6, el lector habrá notado que la estrategia general de la prueba $F$ es la siguiente: hay un modelo más grande, el *modelo no restringido* (8.6.15), y un modelo más pequeño, el *modelo restringido* o *limitado*, que se obtuvo del modelo más grande al eliminar algunas de sus variables, por ejemplo (8.6.18), o al colocar algunas restricciones lineales sobre uno o más coeficientes del modelo más grande, por ejemplo (8.6.16) u (8.6.17).

[14] Si se utiliza el método de máxima verosimilitud en la estimación, entonces hay una prueba semejante a la que estudiará en seguida, a saber, la **prueba de razón de verosimilitud**, la cual es un poco complicada y por tanto se estudia en el apéndice de este capítulo. Para un análisis más a fondo, véase Theil, *op. cit.*, pp. 179-184.

Ajustamos entonces los modelos no restringido y restringido a los datos y obtenemos los coeficientes de determinación respectivos, a saber, $R^2_{\text{NR}}$ y $R^2_{\text{R}}$. Se observan los gl en el modelo no restringido ($= n - k$) y también los gl en el modelo restringido ($= m$), con $m$ como el número de restricciones lineales [por ejemplo 1 en (8.6.16) o en (8.6.18)] o el número de regresoras omitidos del modelo [por ejemplo $m = 4$ si se cumple (8.6.18), pues se supone que hay cuatro regresoras ausentes del modelo]. Así podemos calcular la razón $F$ como se indica en (8.6.9) u (8.6.10), y utilizar esta *regla de decisión: si la F calculada excede $F_\alpha(m, n - k)$, donde $F_\alpha(m, n - k)$ es el F crítico en el nivel de significancia α, se rechaza la hipótesis nula; de lo contrario no se rechaza.*

Ilustremos lo anterior:

**EJEMPLO 8.4**
*Demanda de carne de pollo en Estados Unidos, 1960-1982*

En el ejercicio 7.19, entre otras cosas, se le pidió considerar la siguiente función de demanda de pollos:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_i \qquad \textbf{(8.6.19)}$$

donde $Y$ = consumo de pollo *per cápita*, lbs; $X_2$ = ingreso real disponible *per cápita*, \$; $X_3$ = precio real al menudeo del pollo por lb, ¢; $X_4$ = precio real al menudeo del cerdo por lb, ¢ y $X_5$ = precio real de la carne de res por lb, ¢.

En este modelo $\beta_2$, $\beta_3$, $\beta_4$ y $\beta_5$ son las elasticidades ingreso, precio-propio, precio-cruzado (cerdo) y precio-cruzado (carne de res). (¿Por qué?) De acuerdo con la teoría económica,

$$\begin{aligned}
&\beta_2 > 0 \\
&\beta_3 < 0 \\
&\beta_4 > 0, && \text{si el pollo y el cerdo son productos que compiten} \\
&\quad < 0, && \text{si el pollo y el cerdo son productos complementarios} \\
&\quad = 0, && \text{si el pollo y el cerdo son productos no relacionados} \\
&\beta_5 > 0, && \text{si el pollo y la carne de res son productos que compiten} \\
&\quad < 0, && \text{si el pollo y el cerdo son productos complementarios} \\
&\quad = 0, && \text{si el pollo y el cerdo son productos no relacionados}
\end{aligned} \qquad \textbf{(8.6.20)}$$

Suponga que alguien afirma que el pollo, el cerdo y la carne de res son productos no relacionados en el sentido de que al consumo de pollo no le afectan los precios del cerdo ni de la carne de res. En resumen,

$$H_0: \beta_4 = \beta_5 = 0 \qquad \textbf{(8.6.21)}$$

Por consiguiente, la regresión restringida se transforma en

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t \qquad \textbf{(8.6.22)}$$

La ecuación (8.6.19) es, a todas luces, la regresión no restringida.

Con la información del ejercicio 7.19 obtenemos lo siguiente:

*Regresión no restringida*

$$\begin{aligned}
\widehat{\ln Y_t} = \; & 2.1898 + 0.3425 \ln X_{2t} - 0.5046 \ln X_{3t} + 0.1485 \ln X_{4t} + 0.0911 \ln X_{5t} \\
& (0.1557) \quad (0.0833) \qquad (0.1109) \qquad (0.0997) \qquad (0.1007) \\
& R^2_{\text{NR}} \quad 0.9823
\end{aligned} \qquad \textbf{(8.6.23)}$$

*Regresión restringida*

$$\begin{aligned}
\widehat{\ln Y_t} = \; & 2.0328 + 0.4515 \ln X_{2t} - 0.3772 \ln X_{3t} \\
& (0.1162) \quad (0.0247) \qquad (0.0635) \\
& R^2_{\text{R}} = 0.9801
\end{aligned} \qquad \textbf{(8.6.24)}$$

(*continúa*)

**EJEMPLO 8.4** (*continuación*)

donde las cifras en paréntesis son los errores estándar estimados. *Nota:* Los valores de $R^2$ de (8.6.23) y (8.6.24) son comparables, pues la variable dependiente en los dos modelos es la misma.

Ahora la razón $F$ para probar la hipótesis (8.6.21) es

$$F = \frac{\left(R^2_{\text{NR}} - R^2_{\text{R}}\right)/m}{\left(1 - R^2_{\text{NR}}\right)/(n-k)} \tag{8.6.10}$$

En este caso, el valor de $m$ es 2, pues hay dos restricciones: $\beta_4 = 0$ y $\beta_5 = 0$. Los gl del denominador $(n - k)$ son 18, porque $n = 23$ y $k = 5$ (5 coeficientes $\beta$).

Por consiguiente, la razón $F$ es

$$\begin{aligned} F &= \frac{(0.9823 - 0.9801)/2}{(1 - 0.9823)/18} \\ &= 1.1224 \end{aligned} \tag{8.6.25}$$

que tiene una distribución $F$ con 2 y 18 gl.

En el nivel de 5% se aprecia con claridad que este valor $F$ no es estadísticamente significativo $[F_{0.05}(2,18) = 3.55]$. El valor $p$ es 0.3472. Por consiguiente, no hay razón para rechazar la hipótesis nula: la demanda de pollo no depende de los precios del cerdo ni de la carne de res. En resumen, aceptamos la regresión restringida (8.6.24) como la que representa la función de demanda de pollo.

Observe que la función de demanda satisface las expectativas económicas *a priori,* pues la elasticidad precio-propio es negativa y la elasticidad ingreso es positiva. Sin embargo, la elasticidad precio estimada, en valor absoluto, es estadísticamente menor que la unidad, lo que implica que la demanda de pollo es inelástica al precio. (¿Por qué?) Además, la elasticidad ingreso, si bien es positiva, estadísticamente también es menor que la unidad, lo que sugiere que el pollo no es un artículo de lujo; por convención, se dice que un artículo es de lujo si su elasticidad ingreso es mayor que uno.

## 8.7 Prueba para la estabilidad estructural o paramétrica de los modelos de regresión: la prueba de Chow

Cuando utilizamos un modelo de regresión que implica series de tiempo, tal vez se dé un **cambio estructural** en la relación entre la regresada $Y$ y las regresoras. Por cambio estructural nos referimos a que los valores de los parámetros del modelo no permanecen constantes a lo largo de todo el periodo. A veces el cambio estructural se debe a fuerzas externas (por ejemplo, el embargo petrolero impuesto por la OPEP en 1973 y 1979, o la Guerra del Golfo de 1990-1991), o a cambios en las políticas (como la transformación de un sistema de tasa de cambio fija por otro sistema flexible, alrededor de 1973) o por acciones tomadas por el Congreso estadounidense (por ejemplo, los cambios impositivos iniciados por el presidente Reagan durante sus dos periodos de gobierno, o por los cambios en los salarios mínimos), u otras causas diversas.

¿Cómo saber que de verdad ocurrió un cambio estructural? Para ser específicos, considere los datos de la tabla 8.9 sobre el ingreso personal disponible y el ahorro personales, en miles de millones de dólares, para Estados Unidos de 1970 a 1995. Suponga que deseamos estimar una sencilla función ahorro que relacione el ahorro ($Y$) con el ingreso personal disponible IPD ($X$). Como tenemos los datos se puede calcular una regresión por MCO de $Y$ sobre $X$. Pero, al hacerlo, afirmamos que la relación entre ahorros e IPD no cambió mucho durante el lapso de 26 años. Éste tal vez sea un supuesto difícil de creer. Por ejemplo, se sabe muy bien que en 1982 Estados Unidos experimentó su peor recesión en tiempos de paz. La tasa de desempleo civil alcanzó 9.7%,

**TABLA 8.9**
**Ahorro e ingreso personal disponible (en miles de millones de dólares) para Estados Unidos, 1970-1995**

Fuente: *Economic Report of the President*, 1997, tabla B-28, p. 332.

| Observación | Ahorro | Ingreso | Observación | Ahorro | Ingreso |
|---|---|---|---|---|---|
| 1970 | 61.0 | 727.1 | 1983 | 167.0 | 2 522.4 |
| 1971 | 68.6 | 790.2 | 1984 | 235.7 | 2 810.0 |
| 1972 | 63.6 | 855.3 | 1985 | 206.2 | 3 002.0 |
| 1973 | 89.6 | 965.0 | 1986 | 196.5 | 3 187.6 |
| 1974 | 97.6 | 1 054.2 | 1987 | 168.4 | 3 363.1 |
| 1975 | 104.4 | 1 159.2 | 1988 | 189.1 | 3 640.8 |
| 1976 | 96.4 | 1 273.0 | 1989 | 187.8 | 3 894.5 |
| 1977 | 92.5 | 1 401.4 | 1990 | 208.7 | 4 166.8 |
| 1978 | 112.6 | 1 580.1 | 1991 | 246.4 | 4 343.7 |
| 1979 | 130.1 | 1 769.5 | 1992 | 272.6 | 4 613.7 |
| 1980 | 161.8 | 1 973.3 | 1993 | 214.4 | 4 790.2 |
| 1981 | 199.1 | 2 200.2 | 1994 | 189.4 | 5 021.7 |
| 1982 | 205.5 | 2 347.3 | 1995 | 249.3 | 5 320.8 |

la más alta desde 1948. Un suceso como éste pudo perturbar la relación entre el ahorro y el IPD. Para ver si lo anterior sucedió, dividamos la muestra en dos periodos: 1970-1981 y 1982-1995, antes y después de la recesión de 1982.

Ahora tenemos tres posibles regresiones:

Periodo 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \qquad n_1 = 12$ **(8.7.1)**

Periodo 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \qquad n_2 = 14$ **(8.7.2)**

Periodo 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_t \qquad n = (n_1 + n_2) = 26$ **(8.7.3)**

La regresión (8.7.3) supone que no hay diferencia entre los dos periodos, y por tanto estima la relación entre el ahorro y el IPD para la totalidad del periodo, que consiste en 26 observaciones. En otras palabras, esta regresión supone que el intercepto, así como el coeficiente de la pendiente, permanecen constantes durante todo el periodo; es decir, no hay cambio estructural. Si ésta fuera la situación, entonces $\alpha_1 = \lambda_1 = \gamma_1$ y $\alpha_2 = \lambda_2 = \gamma_2$.

Las regresiones (8.7.1) y (8.7.2) suponen que las regresiones en ambos periodos son distintas; es decir, el intercepto y los coeficientes de las pendientes son diferentes, como se indica mediante los parámetros con subíndice. En las regresiones anteriores, las $u$ representan los términos de error y las $n$ el número de observaciones.

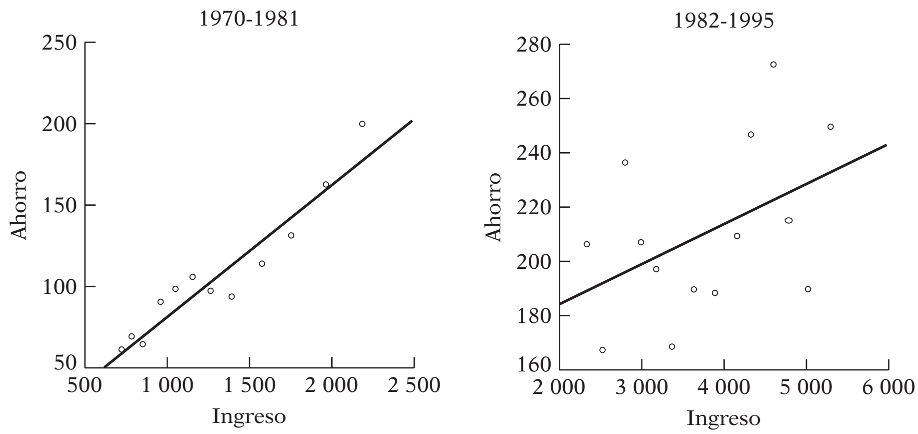
Para los datos de la tabla 8.9, las contrapartes empíricas de las tres regresiones anteriores son las siguientes:

$$\begin{aligned} \hat{Y}_t &= \ 1.0161 \ + \ 0.0803\,X_t \\ t &= (0.0873) \quad (9.6015) \\ R^2 &= 0.9021 \qquad \text{SCR}_1 = 1\,785.032 \qquad \text{gl} = 10 \end{aligned} \tag{8.7.1a}$$

$$\begin{aligned} \hat{Y}_t &= 153.4947 \ + \ 0.0148X_t \\ t &= \ (4.6922) \quad (1.7707) \\ R^2 &= 0.2971 \qquad \text{SCR}_2 = 10\,005.22 \qquad \text{gl} = 12 \end{aligned} \tag{8.7.2a}$$

$$\begin{aligned} \hat{Y}_t &= 62.4226 \ + \ 0.0376\,X_t + \cdots \\ t &= (4.8917) \quad (8.8937) + \cdots \\ R^2 &= 0.7672 \qquad \text{SCR}_3 = 23\,248.30 \qquad \text{gl} = 24 \end{aligned} \tag{8.7.3a}$$

FIGURA 8.3




En las regresiones anteriores, SCR denota la suma de cuadrados residual, y las cifras entre paréntesis son los valores estimados *t*.

Una mirada a las regresiones estimadas indica que la relación entre el ahorro y el IPD no es la misma en los dos subperiodos. La pendiente en las regresiones anteriores de ahorro-ingreso representa la **propensión marginal a ahorrar (PMA)**; es decir, el cambio (medio) en el ahorro como resultado del incremento de un dólar en el ingreso personal disponible. En el periodo 1970-1981 la PMA fue de casi 0.08, en tanto que para el periodo 1982-1995 fue de casi 0.02. Resulta difícil decir si este cambio se debió a las políticas económicas del presidente Reagan. Lo anterior indica, además, que quizá no sea adecuada la **regresión conjunta** (8.7.3*a*), es decir, la que reúne las 26 observaciones y efectúa una regresión común sin tener en cuenta las posibles diferencias en los dos subperiodos. Por supuesto, las afirmaciones anteriores necesitan justificarse por prueba(s) estadística(s) apropiada(s). A propósito, los diagramas de dispersión y las líneas de regresión estimadas se muestran en la figura 8.3.

Ahora bien, las posibles diferencias —es decir, los cambios estructurales— quizá se deban a diferencias en el intercepto o en el coeficiente de la pendiente, o a ambos. ¿Cómo saberlo? De la figura 8.3 se obtiene una intuición gráfica. Pero resulta útil contar con una prueba formal.

Aquí es donde la **prueba de Chow** muestra su valía.[15] La prueba supone que:

1. $u_{1t} \sim N(0, \sigma^2)$ y $u_{2t} \sim N(0, \sigma^2)$. Es decir, los términos de error en las regresiones de los subperiodos están normalmente distribuidos con la misma varianza (homoscedástica) $\sigma^2$.
2. Los dos términos de error ($u_{1t}$ y $u_{2t}$) están independientemente distribuidos.

Los mecanismos de la prueba de Chow son los siguientes:

1. Se estima la regresión (8.7.3), que resulta apropiada si no hay inestabilidad en los parámetros, y se obtiene $\text{SCR}_3$ con gl $= (n_1 + n_2 - k)$, donde $k$ es el número de parámetros estimado, 2 en este caso. Para el ejemplo, $\text{SCR}_3 = 23\ 248.30$. Se llama a $\text{SCR}_3$ la **suma de cuadrados residual restringida ($\text{SCR}_\text{R}$)**, pues se obtiene al imponer las restricciones que $\lambda_1 = \gamma_1$ y $\lambda_2 = \gamma_2$; es decir, las regresiones de los subperiodos no son diferentes.

2. Estime (8.7.1) y obtenga su suma de residuos al cuadrado, $\text{SCR}_1$, con gl $= (n_1 - k)$. En el ejemplo, $\text{SCR}_1 = 1\ 785.032$ y gl $= 10$.

3. Estime (8.7.2) y obtenga su suma de residuos al cuadrado, $\text{SCR}_2$, con gl $= (n_2 - k)$. En el ejemplo, $\text{SCR}_2 = 10\ 005.22$ y gl $= 12$.

[15] Gregory C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, vol. 28, núm. 3, 1960, pp. 591-605.

4. Como los dos conjuntos de muestras se consideran independientes, se pueden sumar $SCR_1$ y $SCR_2$ para obtener lo que se puede llamar **suma de cuadrados residual no restringida ($SCR_{NR}$)**, es decir:

$$SCR_{NR} = SCR_1 + SCR_2 \qquad \text{con gl} = (n_1 + n_2 - 2k)$$

En el caso presente,

$$SCR_{NR} = (1\ 785.032 + 10\ 005.22) = 11\ 790.252$$

5. Ahora bien, la idea que fundamenta la prueba de Chow es que si de hecho no existe un cambio estructural [es decir, las regresiones (8.7.1) y (8.7.2) son esencialmente las mismas], $SCR_R$ y $SCR_{NR}$ no deben ser estadísticamente diferentes. Por tanto, si formamos la siguiente razón:

$$F = \frac{(SCR_R - SCR_{NR})/k}{(SCR_{NR})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]} \tag{8.7.4}$$

Chow demostró que, según la hipótesis nula, las regresiones (8.7.1) y (8.7.2) son (estadísticamente) iguales (es decir, no hay cambios estructurales ni rupturas), así que la razón $F$ dada antes sigue una distribución $F$ con $k$ y $(n_1 + n_2 - 2k)$ gl en el numerador y denominador, respectivamente.

6. Por tanto, no rechazamos la hipótesis nula de la *estabilidad paramétrica* (es decir, no hay cambio estructural) si el valor $F$ calculado en la aplicación no excede el valor crítico $F$ obtenido de la tabla $F$ en el nivel elegido de significancia (o el valor $p$). En este caso se justificaría la regresión conjunta (¿restringida?) (8.7.3). Por el contrario, si el valor $F$ calculado excede el valor crítico $F$, rechazamos la hipótesis de la estabilidad paramétrica y concluimos que las regresiones (8.7.1) y (8.7.2) son diferentes, en cuyo caso la regresión conjunta (8.7.3) es de dudoso valor, por decir lo menos.

De regreso al ejemplo, tenemos que

$$\begin{aligned} F &= \frac{(23\ 248.30 - 11\ 790.252)/2}{(11\ 790.252)/22} \\ &= 10.69 \end{aligned} \tag{8.7.5}$$

De las tablas $F$ vemos que para 2 y 22 gl el valor crítico $F$ a 1% es de 7.72. Por tanto, la probabilidad de obtener un valor $F$ igual o mayor que 10.69 es mucho menor que 1%; de hecho, el valor $p$ sólo es 0.00057.

La prueba de Chow, por tanto, parece apoyar la anterior conjetura de que la relación ahorro-ingreso sufrió un cambio estructural en Estados Unidos en el periodo 1970-1995, pues da por hecho que se satisfacen las suposiciones subyacentes en la prueba. Diremos más al respecto muy pronto.

A propósito, observe que la prueba de Chow se generaliza con facilidad para abarcar casos de más de una ruptura estructural. Por ejemplo, si pensamos que la relación ahorro-ingreso cambió después de que el presidente Clinton asumió el cargo en enero de 1992, dividimos la muestra en tres periodos: 1970-1981, 1982-1991 y 1992-1995, y aplicamos la prueba de Chow. Por supuesto, se tendrán cuatro términos SCR, uno para cada subperiodo y otro para todos los datos juntos. Pero la lógica de la prueba sigue siendo la misma. Ahora disponemos de los datos para 2007, por lo que se puede extender el último periodo hasta ese año.

Se deben tener presente algunas advertencias respecto de la prueba de Chow:

1. Las suposiciones en la prueba deben satisfacerse. Por ejemplo, se debe averiguar si las varianzas de los errores en las regresiones (8.7.1) y (8.7.2) son las mismas. Analizaremos pronto este punto.

2. La prueba de Chow dirá sólo si las dos regresiones (8.7.1) y (8.7.2) son diferentes, pero no señalará si la diferencia se debe a los interceptos o a las pendientes, o a ambos. Pero en el capítulo 9, sobre variables dicótomas, veremos cómo responder lo anterior.

3. La prueba de Chow supone que se conoce(n) el(los) punto(s) de ruptura estructural. En el ejemplo anterior supusimos que se dio en 1982. No obstante, si no es posible determinar cuándo ocurrió en verdad el cambio estructural, habrá que utilizar otros métodos.[16]

Antes de finalizar con el tema de la prueba de Chow y la regresión ahorro-ingreso, examinaremos una de las suposiciones implícitas en dicha prueba, a saber, que las varianzas de los errores en los dos periodos son las mismas. Como no pueden observarse las verdaderas varianzas de los errores se pueden obtener sus estimaciones a partir de las SCR dadas en las regresiones (8.7.1*a*) y (8.7.2*a*), a saber:

$$\hat{\sigma}_1^2 = \frac{\text{SCR}_1}{n_1 - 2} = \frac{1\,785.032}{10} = 178.5032 \tag{8.7.6}$$

$$\hat{\sigma}_2^2 = \frac{\text{SCR}_2}{n_2 - 2} = \frac{10\,005.22}{14 - 2} = 833.7683 \tag{8.7.7}$$

Observe que, como hay dos parámetros estimados en cada ecuación, se resta 2 al número de observaciones para obtener los gl. Según los supuestos en la prueba de Chow, $\hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ son estimadores insesgados de las verdaderas varianzas de los dos subperiodos. Como resultado, se puede probar que si $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$, es decir, las varianzas en las dos subpoblaciones son las mismas (como supuso la prueba de Chow), entonces se demuestra que

$$\frac{\left(\hat{\sigma}_1^2/\sigma_1^2\right)}{\left(\hat{\sigma}_2^2/\sigma_2^2\right)} \sim F_{(n_1-k),(n_2-k)} \tag{8.7.8}$$

sigue una distribución $F$ con $(n_1 - k)$ y $(n_2 - k)$ gl en el numerador y el denominador, respectivamente (para este ejemplo, $k = 2$), porque sólo hay dos parámetros en cada subregresión.

Por supuesto, si $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$, la prueba $F$ anterior se reduce a calcular

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \tag{8.7.9}$$

*Nota:* Por convención, se coloca la más grande de las varianzas estimadas en el numerador. (Véase el **apéndice A** para detalles sobre la distribución $F$ y otras distribuciones de probabilidad.)

Si calculamos esta $F$ en una aplicación y la comparamos con el valor crítico $F$ que tenga los apropiados gl, podemos decidir rechazar o no la hipótesis nula de que las varianzas en las dos subpoblaciones son las mismas. Si no rechazamos la hipótesis nula se puede utilizar la prueba de Chow.

De nuevo en la regresión ahorro-ingresos, obtenemos el siguiente resultado:

$$F = \frac{833.7683}{178.5032} = 4.6701 \tag{8.7.10}$$

Según la hipótesis nula de la igualdad de las varianzas en las dos subpoblaciones, este valor $F$ sigue la distribución $F$ con 12 y 10 gl en el numerador y denominador, respectivamente. (*Nota:* Se puso la mayor de las dos varianzas estimadas en el numerador.) De las tablas $F$ del **apéndice D** se ve que los valores críticos $F$ de 5 y 1% para 12 y 10 gl son 2.91 y 4.71, respectivamente. El

[16] Para un análisis detallado, véase William H. Greene, *Econometric Analysis*, 4a. ed., Prentice Hall, Englewood Cliffs, Nueva Jersey, 2000, pp. 293-297.

valor calculado $F$ es significativo en el nivel de 5% y es casi significativo en el de 1%. Por tanto, la conclusión sería que las varianzas de las dos subpoblaciones no son iguales y, por tanto, en sentido estricto, no debemos utilizar la prueba de Chow.

Aquí el propósito fue presentar el procedimiento de la prueba de Chow más común en el trabajo aplicado. Si las varianzas de los errores en las dos subpoblaciones son heteroscedásticas, se puede modificar la prueba de Chow. Pero el procedimiento trasciende el alcance de este libro.[17]

Otro punto en el que ya hicimos hincapié fue en que la prueba de Chow es sensible a la elección del tiempo en que los parámetros de la regresión deben cambiarse. En este ejemplo, supusimos que el cambio tal vez tuvo lugar en la recesión de 1982. Si hubiésemos supuesto que fue en 1981, cuando Ronald Reagan comenzó su mandato, se descubriría que el valor $F$ calculado es diferente. De hecho, en el ejercicio 8.34 se pide al lector verificar esto.

Si no deseamos elegir el punto en el que pudo ocurrir la ruptura de la relación subyacente, elegimos otros métodos, como la **prueba residual recursiva**. Analizaremos este tema en el capítulo 13, que trata sobre el análisis de especificación del modelo.

## 8.8 Predicción con regresión múltiple

En la sección 5.10 mostramos que el modelo estimado de regresión con dos variables sirve para 1) *predicción de la media*, es decir, predicción puntual sobre la función de regresión poblacional (FRP), y también para 2) *predicción individual*, es decir, predicción de un valor individual de $Y$, dado el valor de la regresora $X = X_0$, donde $X_0$ es el valor numérico específico de $X$.

La regresión múltiple estimada también sirve para fines similares, y el procedimiento para hacerlo es una extensión directa del caso de dos variables, con excepción de las fórmulas para estimar la varianza y el error estándar de los valores de pronóstico [comparables a (5.10.2) y (5.10.6) del modelo con dos variables], las cuales son más bien complejas y se manejan mejor mediante los métodos matriciales estudiados en el **apéndice C**. Por supuesto, la mayoría del software de regresión efectúa esto de manera rutinaria, así que no hay necesidad de recurrir a la formulación matricial, la cual se da en el **apéndice C** para beneficio de los estudiantes con inclinaciones matemáticas. Ahí también se proporciona un ejemplo completo.

## *8.9 La tríada de las pruebas de hipótesis: razón de verosimilitud (RV), de Wald (W) y del multiplicador de Lagrange (ML)[18]

En este capítulo y en los anteriores utilizamos, generalmente, las pruebas $t$, $F$ y ji cuadrada para probar diversas hipótesis en el contexto de los modelos de regresión lineal (en parámetros). Pero una vez que salimos del cómodo mundo de los modelos de regresión lineal, se necesitan métodos para probar hipótesis con los que se manejen modelos de regresión, lineales o no lineales.

Con la conocida tríada de **pruebas de verosimilitud, de Wald y del multiplicador de Lagrange** se logra este propósito. Lo interesante es que *asintóticamente* (es decir, en muestras

* Opcional.

[17] Para un estudio de la prueba de Chow con heteroscedasticidad, véase William H. Greene, *Econometric Analysis*, 4a. ed., Prentice Hall, Englewood Cliffs, Nueva Jersey, 2000, pp. 292-293, y Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, Reino Unido, 1994, p. 51.

[18] Véase una exposición sencilla en A. Buse, "The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note", *American Statistician*, vol. 36, 1982, pp. 153-157.

grandes) las tres pruebas son equivalentes en cuanto a que la estadística de prueba asociada a cada prueba sigue la distribución ji cuadrada.

Aunque estudiaremos la **prueba de la razón de verosimilitud** en el apéndice de este capítulo, en general no utilizaremos este tipo de pruebas en este libro de texto por la razón pragmática de que en muestras pequeñas o finitas, que son las que, por desgracia, maneja la mayoría de los investigadores, bastará la prueba $F$ que hemos visto hasta ahora. Como lo anotan Davidson y MacKinnon:

> Para modelos de regresión lineal, con errores normales o sin ellos, no hay necesidad de revisar ML, W y RV, pues, al hacerlo no se gana información adicional a la contenida en $F$.[19]

## *8.10 Prueba de la forma funcional de la regresión: elección entre modelos de regresión lineal y log-lineal

La selección entre un modelo de regresión lineal (la regresora es una función lineal de las regresoras) o un modelo de regresión log-lineal (el logaritmo de la regresora es función de los logaritmos de las regresoras) es la eterna pregunta en el análisis empírico. Se puede utilizar una prueba propuesta por MacKinnon, White y Davidson, que se denomina, por brevedad, **prueba MWD**, para escoger entre los dos modelos.[20]

Para ilustrar esta prueba suponga lo siguiente:

$H_0$: *Modelo lineal*: $Y$ es una función lineal de las regresoras, las $X$.

$H_1$: *Modelo log-lineal*: ln $Y$ es función lineal de los logaritmos de las regresoras, los logaritmos de las $X$.

donde, como es usual, $H_0$ y $H_1$ denotan las hipótesis nula y alterna.

La prueba MWD comprende los siguientes pasos:[21]

**Paso I:** Estime el modelo lineal y obtenga los valores $Y$ estimados. Llámelos $Yf$ (es decir, $\hat{Y}$).

**Paso II:** Estime el modelo log-lineal y obtenga los valores ln $Y$ estimados; denomine ln $f$ (es decir, $\widehat{\ln Y}$.

**Paso III:** Obtenga $Z_1 = (\ln Yf - \ln f)$.

**Paso IV:** Efectúe la regresión de $Y$ sobre las $X$ y $Z_1$ obtenida en el paso III. Rechace $H_0$ si el coeficiente de $Z_1$ es estadísticamente significativo mediante la prueba $t$ usual.

**Paso V:** Obtenga $Z_2 = (\text{antilog de } \ln f - Yf)$.

**Paso VI:** Efectúe la regresión del logaritmo de $Y$ sobre los logaritmos de las $X$ y $Z_2$. Rechace $H_1$ si el coeficiente de $Z_2$ es estadísticamente significativo mediante la prueba $t$ usual.

Aunque la prueba MWD parece compleja, su lógica es muy simple. Si el modelo lineal es en realidad el modelo correcto, la variable construida $Z_1$ no debe ser estadísticamente significativa en el paso IV, pues en ese caso los valores $Y$ estimados del modelo lineal y los estimados del modelo log-lineal (después de obtener sus valores antilog para efectos comparativos) no deben diferir. El mismo comentario vale para la hipótesis alterna $H_1$.

---

* Opcional.

[19] Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, p. 456.

[20] J. Mackinnon, H. White y R. Davidson, "Tests for Model Specification in the Presence of Alternative Hypothesis: Some Further Results", *Journal of Econometrics*, vol. 21, 1983, pp. 53-70. Se propone una prueba similar en A. K. Bera y C. M. Jarque, "Model Specification Tests: A Simultaneous Approach", *Journal of Econometrics*, vol. 20, 1982, pp. 59-82.

[21] Este análisis se basa en William H. Greene, *ET: The Econometrics Toolkit Version 3*, Econometric Software, Bellport, Nueva York, 1992, pp. 245-246.

**EJEMPLO 8.5**
*Demanda de rosas*

Consulte el ejercicio 7.16, en el cual se presenta información sobre la demanda de rosas en el área metropolitana de Detroit de 1971-III a 1975-II. Para fines ilustrativos consideraremos la demanda de rosas como función sólo de los precios de las rosas y de los claveles, y dejaremos fuera, por el momento, la variable ingreso. Ahora consideremos los siguientes modelos:

Modelo lineal: $$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_t \tag{8.10.1}$$

Modelo log-lineal: $$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t \tag{8.10.2}$$

donde $Y$ es la cantidad de rosas por docenas, $X_2$ es el precio promedio de las rosas al mayoreo ($/docena) y $X_3$ es el precio promedio de los claveles al mayoreo ($/docena). Se espera, *a priori*, que $\alpha_2$ y $\beta_2$ sean negativos (¿por qué?), y que $\alpha_3$ y $\beta_3$ sean positivos (¿por qué?). Como se sabe, los coeficientes de pendiente en los modelos log-lineal son coeficientes de elasticidad.

Los resultados de las regresiones son los siguientes:

$$\begin{aligned}\hat{Y}_t &= 9\,734.2176 - 3\,782.1956X_{2t} + 2\,815.2515X_{3t}\\ t &= \quad (3.3705) \qquad (-6.6069) \qquad\qquad (2.9712)\\ &\qquad F = 21.84 \qquad R^2 = 0.77096\end{aligned} \tag{8.10.3}$$

$$\begin{aligned}\widehat{\ln Y_t} &= 9.2278 - 1.7607 \ln X_{2t} + 1.3398 \ln X_{3t}\\ t &= (16.2349) \quad (-5.9044) \qquad (2.5407)\\ &\qquad F = 17.50 \qquad R^2 = 0.7292\end{aligned} \tag{8.10.4}$$

Como lo indican estos resultados, ambos modelos, el lineal y el log-lineal, parecen ajustarse a la información razonablemente bien: los parámetros tienen los signos esperados y los valores $t$ y $R^2$ son estadísticamente significativos.

Para decidir entre estos modelos con base en la **prueba MWD**, se prueba primero la hipótesis de que el modelo verdadero es lineal. Luego, según el paso IV de la prueba, obtenemos la siguiente regresión:

$$\begin{aligned}\hat{Y}_t &= 9\,727.5685 - 3\,783.0623X_{2t} + 2\,817.7157X_{3t} + 85.2319Z_{1t}\\ t &= \quad (3.2178) \qquad (-6.3337) \qquad\qquad (2.8366) \qquad (0.0207)\\ &\qquad F = 13.44 \qquad R^2 = 0.7707\end{aligned} \tag{8.10.5}$$

Como el coeficiente de $Z_1$ no es estadísticamente significativo (el valor $p$ del $t$ estimado es 0.98), no rechazamos la hipótesis de que el verdadero modelo es lineal.

Suponga que aceleramos el proceso y que el verdadero modelo es log-lineal. Según el paso VI de la prueba MWD, obtenemos los siguientes resultados de la regresión:

$$\begin{aligned}\widehat{\ln Y_t} &= 9.1486 - 1.9699 \ln X_t + 1.5891 \ln X_{2t} - 0.0013Z_{2t}\\ t &= (17.0825) \quad (-6.4189) \qquad (3.0728) \qquad (-1.6612)\\ &\qquad F = 14.17 \qquad R^2 = 0.7798\end{aligned} \tag{8.10.6}$$

El coeficiente de $Z_2$ es estadísticamente significativo en un nivel de significancia de 12% (el valor $p$ es 0.1225). Por consiguiente, podemos rechazar la hipótesis de que el verdadero modelo es log-lineal en este nivel de significancia. Por supuesto, si utilizamos los niveles de significancia convencionales de 1 y 5%, entonces no podemos rechazar la hipótesis de que el verdadero modelo es log-lineal. Como muestra este ejemplo, es muy posible que en una situación dada no podamos rechazar una u otra de las especificaciones.

## Resumen y conclusiones

1. En este capítulo se extendieron y refinaron las ideas sobre estimación por intervalos y pruebas de hipótesis presentadas en el capítulo 5 en el contexto del modelo de regresión lineal con dos variables.
2. En la regresión múltiple, la prueba de *significancia individual* de un coeficiente de regresión parcial (con la prueba *t*) y la prueba de la *significancia general* de la regresión (es decir, $H_0$: todos los coeficientes de pendiente parcial son cero o $R^2 = 0$) no son la misma cosa.
3. En particular, encontrar que uno o más coeficientes de regresión parcial no son estadísticamente significativos con base en la prueba *t individual* no significa que todos los coeficientes de regresión parcial (colectivamente) tampoco sean significativos. Esta última hipótesis sólo se prueba mediante la prueba *F*.
4. La **prueba *F*** tiene gran versatilidad, pues con ella se pueden probar diversas hipótesis, como 1) si un coeficiente de regresión individual es estadísticamente significativo, 2) si todos los coeficientes de pendiente parciales son cero, 3) si dos o más coeficientes son estadísticamente iguales, 4) si los coeficientes satisfacen algunas restricciones lineales y 5) si el modelo de regresión posee estabilidad estructural.
5. Como en el caso de dos variables, el modelo de regresión múltiple sirve para fines de predicción de media y/o individual.

## EJERCICIOS

### Preguntas

8.1. Suponga que desea estudiar el comportamiento de las ventas de un producto, por ejemplo, automóviles, sobre un número de años, y suponga que alguien sugiere ensayar los siguientes modelos:

$$Y_t = \beta_0 + \beta_1 t$$
$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

donde $Y_t$ = ventas en el tiempo *t* y *t* = tiempo, medido en años. El primer modelo postula que la variable ventas es una función lineal del tiempo, mientras que el segundo plantea que es función cuadrática del tiempo.

*a*) Analice las propiedades de estos modelos.

*b*) ¿Cómo decidiría entre los dos modelos?

*c*) ¿En qué situaciones sería útil el modelo cuadrático?

*d*) Obtenga información sobre ventas de automóviles en Estados Unidos durante los pasados 20 años y vea qué modelo se ajusta mejor a los datos.

8.2. Demuestre que la razón *F* de (8.4.16) es igual a la razón *F* de (8.4.18). (*Sugerencia:* SCE/SCT = $R^2$.)

8.3. Demuestre que las pruebas *F* de (8.4.18) y (8.6.10) son equivalentes.

8.4. Justifique los supuestos en (8.6.11) y (8.6.12).

8.5. Considere la función de producción Cobb-Douglas

$$Y = \beta_1 L^{\beta_2} K^{\beta_3} \quad (1)$$

donde $Y$ = producción, $L$ = insumo trabajo y $K$ = insumo capital. Al dividir (1) entre $K$, obtenemos:

$$(Y/K) = \beta_1 (L/K)^{\beta_2} K^{\beta_2+\beta_3-1} \quad (2)$$

Al tomar el logaritmo natural de (2) tenemos:

$$\ln(Y/K) = \beta_0 + \beta_2 \ln(L/K) + (\beta_2 + \beta_3 - 1)\ln K + u_i \quad (3)$$

donde $\beta_0 = \ln \beta_1$.

*a*) Suponga que tenía información para hacer la regresión (3). ¿Cómo probaría la hipótesis de que hay rendimientos constantes a escala, es decir, $(\beta_2 + \beta_3) = 1$?

*b*) De existir rendimientos constantes a escala, ¿cómo interpretaría la regresión (3)?

*c*) ¿Hay alguna diferencia si divide (1) entre $L$ en lugar de hacerlo entre $K$?

8.6. **Valores críticos de $R^2$ cuando la verdadera $R^2 = 0$.** La ecuación (8.4.11) da la relación entre $F$ y $R^2$ según la hipótesis de que todos los coeficientes de pendiente parciales son simultáneamente iguales a cero (es decir, $R^2 = 0$). De la misma manera en que podemos encontrar el valor $F$ crítico en el nivel de significancia $\alpha$ de la tabla $F$, es posible encontrar el valor $R^2$ crítico a partir de la siguiente relación:

$$R^2 = \frac{(k-1)F}{(k-1)F + (n-k)}$$

donde $k$ es el número de parámetros en el modelo de regresión, inclusive el intercepto, y donde $F$ es el valor $F$ crítico en el nivel de significancia $\alpha$. Si el $R^2$ observado excede el $R^2$ crítico obtenido de la fórmula anterior rechazamos la hipótesis de que el verdadero $R^2$ es cero.

Establezca la fórmula anterior y encuentre el valor $R^2$ crítico (para $\alpha = 5\%$) para la regresión (8.1.4).

8.7. Con base en información anual de los años 1968-1987 se obtuvieron los siguientes resultados de regresión:

$$\hat{Y}_t = -859.92 + 0.6470X_{2t} - 23.195X_{3t} \qquad R^2 = 0.9776 \qquad (1)$$

$$\hat{Y}_t = -261.09 + 0.2452X_{2t} \qquad R^2 = 0.9388 \qquad (2)$$

donde $Y =$ gasto de Estados Unidos en bienes importados, miles de millones de dólares de 1982, $X_2 =$ ingreso personal disponible, miles de millones de dólares de 1982 y $X_3 =$ variable de tendencia. *Cierto o falso*: El error estándar de $X_3$ en (1) es 4.2750. Muestre sus cálculos. (*Sugerencia:* Utilice la relación entre $R^2$, $F$ y $t$.)

8.8. Suponga que en la regresión

$$\ln(Y_i/X_{2i}) = \alpha_1 + \alpha_2 \ln X_{2i} + \alpha_3 \ln X_{3i} + u_i$$

se conocen los valores de los coeficientes de regresión y sus errores estándar.* Así, ¿cómo estimaría los parámetros y los errores estándar del siguiente modelo de regresión?

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

8.9. Suponga lo siguiente:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

donde $Y$ es el gasto de consumo personal, $X_2$ es el ingreso personal y $X_3$ es la riqueza personal.† El término $(X_{2i} X_{3i})$ se conoce como **término de interacción**. ¿Qué significa esta expresión? ¿Cómo probaría la hipótesis de que la propensión marginal a consumir (PMC) (es decir, $\beta_2$) es independiente de la riqueza del consumidor?

* Adaptado de Peter Kennedy, *A Guide to Econometrics*, the MIT Press, 3a. ed., Cambridge, Massachusetts, 1992, p. 310.

† *Ibid.*, p. 327.

8.10. Se le dan los siguientes resultados de regresión:

$$\begin{aligned} \hat{Y}_t &= 16\,899 \quad - 2\,978.5X_{2t} && R^2 = 0.6149 \\ t &= \quad (8.5152) \quad (-4.7280) && \\ \hat{Y}_t &= 9\,734.2 \quad - 3\,782.2X_{2t} + 2\,815X_{3t} && R^2 = 0.7706 \\ t &= \quad (3.3705) \quad (-6.6070) \quad (2.9712) && \end{aligned}$$

¿Puede encontrar el tamaño de la muestra en la cual se basan estos resultados? (*Sugerencia:* Recuerde la relación entre los valores de $R^2$, $F$ y $t$.)

8.11. Con base en el análisis de pruebas de hipótesis *individuales* y *conjuntas* a partir de las pruebas $t$ y $F$, respectivamente, ¿cuáles de las siguientes situaciones son probables?

1. Rechazar la nula conjunta con base en el estadístico $F$, pero no rechazar cada nula por separado con base en las pruebas $t$ individuales.
2. Rechazar la nula conjunta con base en el estadístico $F$, rechazar una hipótesis individual con base en la prueba $t$ y no rechazar otras hipótesis individuales con base en la prueba $t$.
3. Rechazar la nula conjunta con base en el estadístico $F$, y rechazar cada hipótesis nula por separado con base en pruebas $t$ individuales.
4. No rechazar la nula conjunta con base en el estadístico $F$, y no rechazar cada nula por separado con base en las pruebas $t$ individuales.
5. No rechazar la nula conjunta con base en el estadístico $F$, rechazar una hipótesis individual con base en una prueba $t$ y no rechazar otras hipótesis individuales con base en la prueba $t$.
6. No rechazar la nula conjunta con base en el estadístico $F$, pero rechazar cada nula por separado con base en pruebas $t$ individuales.*

## Ejercicios empíricos

8.12. Consulte el ejercicio 7.21.

*a*) ¿Cuáles son las elasticidades del ingreso real y de la tasa de interés de los balances reales de efectivo?

*b*) ¿Son las elasticidades anteriores, consideradas en forma individual, estadísticamente significativas?

*c*) Pruebe la significancia general de la regresión estimada.

*d*) ¿Es la elasticidad del ingreso de la demanda por balances reales de efectivo significativamente diferente de la unidad?

*e*) ¿Debe conservarse en el modelo la variable tasa de interés? ¿Por qué?

8.13. De los datos de 46 estados de Estados Unidos para 1992, Baltagi obtuvo los siguientes resultados de regresión:†

$$\begin{aligned} \widehat{\log C} &= 4.30 - 1.34 \log P + 0.17 \log Y \\ \text{ee} &= (0.91) \quad (0.32) \qquad (0.20) \qquad \bar{R}^2 = 0.27 \end{aligned}$$

donde $C$ = consumo de cigarrillos, paquetes al año
$P$ = precio real por paquete
$Y$ = ingreso disponible real *per cápita*

* Citado de Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading, Massachusetts, 1991, p. 79.

† Véase Badi H. Baltagi, *Econometrics*, Springer-Verlag, Nueva York, 1998, p. 111.

*a*) ¿Cuál es la elasticidad de la demanda para los cigarrillos respecto del precio? ¿Es estadísticamente significativa? Si ése es el caso, ¿es estadísticamente diferente de 1?

*b*) ¿Cuál es la elasticidad del ingreso de la demanda de cigarrillos? ¿Es estadísticamente significativa? Si no es así, ¿cuáles serían las razones?

*c*) ¿Cómo obtendría $R^2$ de la $R^2$ *ajustada* dada antes?

8.14. De una muestra de 209 empresas, Wooldridge obtuvo los siguientes resultados de regresión:*

$$\widehat{\log(\text{salario})} = 4.32 + 0.280 \log(\text{ventas}) + 0.0174\ \text{rc} + 0.00024\ \text{rae}$$
$$\text{ee} = (0.32) \quad (0.035) \quad (0.0041) \quad (0.00054)$$
$$R^2 = 0.283$$

donde salario = salario del director general
ventas = ventas anuales de la empresa
rc = rendimientos del capital, en porcentaje
rae = rendimientos de las acciones de la empresa

y donde las cifras entre paréntesis son los errores estándar estimados.

*a*) Interprete la regresión anterior tomando en cuenta cualesquiera expectativas que tenga respecto de los signos de los distintos coeficientes.

*b*) ¿Cuáles de los coeficientes son estadísticamente significativos en lo *individual*, en el nivel de 5%?

*c*) ¿Cuál es la significancia general de la regresión? ¿Qué pruebas utilizaría? ¿Por qué?

*d*) ¿Puede interpretar los coeficientes rc y rae como coeficientes de elasticidad? ¿Por qué?

8.15. Si se supone que $Y$ y $X_2, X_3, \ldots, X_k$ siguen una distribución normal conjunta y que la hipótesis nula plantea que las correlaciones parciales poblacionales son individualmente iguales a cero, R. A. Fisher demostró que

$$t = \frac{r_{1\,2.3\,4\ldots k}\sqrt{n-k-2}}{\sqrt{1-r^2_{1\,2.3\,4\ldots k}}}$$

sigue la distribución $t$ con $n - k - 2$ gl, donde $k$ es el coeficiente de correlación parcial de grado $k$ y $n$ es el número total de observaciones. (*Nota:* $r_{12.3}$ es un coeficiente de correlación parcial de primer orden, $r_{1\,2.3\,4}$ es un coeficiente de correlación parcial de segundo orden y así sucesivamente.) Consulte el ejercicio 7.2. Si se supone que $Y$ y $X_2$ y $X_3$ siguen una distribución normal conjunta, calcule las tres correlaciones parciales $r_{1\,2.3}$, $r_{1\,3.2}$ y $r_{2\,3.1}$, y pruebe su significancia según la hipótesis de que las correlaciones poblacionales correspondientes son individualmente iguales a cero.

8.16. Al estudiar la demanda de tractores agrícolas en Estados Unidos durante los periodos 1921-1941 y 1948-1957, Griliches† obtuvo los siguientes resultados:

$$\widehat{\log Y_t} = \text{constante} - 0.519 \log X_{2t} - 4.933 \log X_{3t} \qquad R^2 = 0.793$$
$$(0.231) \qquad (0.477)$$

* Véase Jeffrey M. Wooldridge, *Introductory Econometrics*, South-Western Publishing, 2000, pp. 154-155.

† Z. Griliches, "The Demand for a Durable Input: Farm Tractors in the United States, 1921-1957", en *The Demand for Durable Goods*, Arnold C. Harberger (ed.), The University of Chicago Press, Chicago, 1960, tabla 1, p. 192.

donde $Y_t$ = valor de las existencias de tractores en las granjas el 1 de enero, en dólares de 1935-1939, $X_2$ = índice de precios pagado por los tractores dividido entre un índice de precios recibidos por todas las cosechas en el tiempo $t-1$, $X_3$ = tasa de interés prevalente en el año $t-1$, y los errores estándar están dados entre paréntesis.

*a*) Interprete la regresión anterior.

*b*) ¿Son los coeficientes de pendiente estimados estadísticamente significativos de manera individual? ¿Son significativamente diferentes de la unidad?

*c*) Utilice la técnica de análisis de varianza para probar la significancia de la regresión en general. *Sugerencia:* Utilice la variante $R^2$ de la técnica ANOVA.

*d*) ¿Cómo calcularía la elasticidad tasa de interés de la demanda de tractores agrícolas?

*e*) ¿Cómo probaría la significancia del $R^2$ estimado?

8.17. Considere la siguiente ecuación de determinación de salarios para la economía británica* durante el periodo 1950-1969:

$$\begin{aligned}\hat{W}_t = \;& 8.582 + 0.364(\text{PF})_t + 0.004(\text{PF})_{t-1} - 2.560U_t \\ & (1.129) \quad (0.080) \qquad (0.072) \qquad (0.658) \\ & R^2 = 0.873 \qquad \text{gl} = 15\end{aligned}$$

donde $W$ = sueldos y salarios por empleado
PF = precios del producto final al factor de costo
$U$ = desempleo en Gran Bretaña como porcentaje del número total de empleados de Gran Bretaña
$t$ = tiempo

(Las cifras en paréntesis son los errores estándar estimados.)

*a*) Interprete la ecuación anterior.

*b*) ¿Son los coeficientes estimados individualmente significativos?

*c*) ¿Cuál es el razonamiento para introducir $(\text{PF})_{t-1}$?

*d*) ¿Debe eliminarse del modelo la variable $(\text{PF})_{t-1}$? ¿Por qué?

*e*) ¿Cómo calcularía la elasticidad de sueldos y salarios por empleado respecto de la tasa de desempleo U?

8.18. Una variación de la ecuación de determinación de salarios del ejercicio 8.17 es la siguiente:†

$$\begin{aligned}\hat{W}_t = \;& 1.073 + 5.288V_t - 0.116X_t + 0.054M_t + 0.046M_{t-1} \\ & (0.797) \quad (0.812) \quad (0.111) \quad (0.022) \quad (0.019) \\ & R^2 = 0.934 \quad \text{gl} = 14\end{aligned}$$

donde $W$ = sueldos y salarios por empleado
$V$ = empleos vacantes como porcentaje del número total de empleados en Gran Bretaña
$X$ = producto interno bruto por persona empleada
$M$ = precios de importaciones
$M_{t-1}$ = precios de importaciones en el año anterior (o rezagado)

(Los errores estándar estimados están dados entre paréntesis.)

* Tomado de *Prices and Earnings in 1951-1969: An Econometric Assessment*, Departamento de Empleo, HMSO, 1971, ecuación (19), p. 35.

† *Ibid.*, ecuación (67), p. 37.

*a*) Interprete la ecuación anterior.

*b*) ¿Cuáles de los coeficientes estimados son estadísticamente significativos individualmente?

*c*) ¿Cuál es el razonamiento para la introducción de la variable *X*? *A priori*, ¿se espera que el signo de *X* sea negativo?

*d*) ¿Cuál es el propósito de incluir $M_t$ y $M_{t-1}$ en el modelo?

*e*) ¿Cuáles variables pueden sacarse del modelo? ¿Por qué?

*f*) Pruebe la significancia general de la regresión observada.

8.19. Para la función de demanda de pollos estimada en (8.6.24), ¿es la elasticidad ingreso estimada igual a 1? ¿Es la elasticidad precio igual a $-1$?

8.20. Para la función de demanda (8.6.24), ¿cómo probaría la hipótesis de que la elasticidad ingreso es igual en valor pero opuesta en signo a la elasticidad precio de la demanda? Muestre los cálculos necesarios. (*Nota:* cov $\hat{\beta}_2, \hat{\beta}_3] = -0.00142$.)

8.21. Consulte la función de demanda de rosas del ejercicio 7.16. Centre su atención en la especificación logarítmica.

*a*) ¿Cuál es la elasticidad precio-propio de la demanda estimada (es decir, la elasticidad respecto del precio de las rosas)?

*b*) ¿Es estadísticamente significativa?

*c*) De ser así, ¿es significativamente diferente de la unidad?

*d*) *A priori*, ¿cuáles son los signos esperados de $X_3$ (precio de los claveles) y $X_4$ (ingreso)? ¿Corresponden los resultados empíricos a estas expectativas?

*e*) Si los coeficientes de $X_3$ y $X_4$ no son estadísticamente significativos, ¿cuáles pueden ser las razones?

8.22. Consulte el ejercicio 7.17, relacionado con la actividad de exploración de pozos.

*a*) ¿Es cada uno de los coeficientes de pendiente estimados estadísticamente significativo individualmente en el nivel de 5%?

*b*) ¿Rechazaría la hipótesis de que $R^2 = 0$?

*c*) ¿Cuál es la tasa de crecimiento instantánea de la actividad de exploración durante el periodo 1948-1978? ¿Cuál es la tasa de crecimiento compuesta correspondiente?

8.23. Consulte los resultados de la regresión de los gastos de presupuesto de defensa de Estados Unidos estimada en el ejercicio 7.18.

*a*) Comente, en términos generales, los resultados de regresión estimados.

*b*) Prepare la tabla ANOVA y pruebe la hipótesis de que todos los coeficientes de pendiente parciales son iguales a cero.

8.24. La siguiente función se conoce como la **función de producción trascendental** (FPT), generalización de la conocida función de producción Cobb-Douglas:

$$Y_i = \beta_1 L^{\beta_2} k^{\beta_3} e^{\beta_4 L + \beta_5 K}$$

donde $Y$ = producción, $L$ = insumo trabajo y $K$ = insumo capital.

Después de tomar logaritmos y de sumar el término de perturbación estocástico, obtenemos la FPT estocástica como

$$\ln Y_i = \beta_0 + \beta_2 \ln L_i + \beta_3 \ln K_i + \beta_4 L_i + \beta_5 K_i + u_i$$

donde $\beta_0 = \ln \beta_1$.

*a*) ¿Cuáles son las propiedades de esta función?

*b*) Para reducir la FPT a la función de producción Cobb-Douglas, ¿cuáles deben ser los valores de $\beta_4$ y $\beta_5$?

*c*) Si tuviera la información, ¿cómo haría para encontrar la forma en la que la FPT se reduce a la función de producción Cobb-Douglas? ¿Qué procedimiento de prueba utilizaría?

*d*) Verifique si la FPT se ajusta a los datos de la tabla 8.8. Muestre sus cálculos.

8.25. *Precios de energía y formación de capital: Estados Unidos, 1948-1978.* Para probar la hipótesis de que un aumento en el precio de la energía relativo a la producción provoca un descenso en la productividad del capital *existente* y de los recursos laborales, John A. Tatom estimó la siguiente función de producción para Estados Unidos, durante el periodo trimestral 1948-I a 1978-II:*

$$\begin{aligned}\widehat{\ln(y/k)} = \underset{(16.33)}{1.5492} &+ \underset{(21.69)}{0.7135}\ln(h/k) - \underset{(-6.42)}{0.1081}\ln(P_e/P) \\ &+ \underset{(15.86)}{0.0045t} \qquad R^2 = 0.98\end{aligned}$$

donde $y$ = producción real en el sector de negocios privado
$k$ = una medida del flujo de servicios de capital
$h$ = horas por persona en el sector de negocios privado
$P_e$ = índice de precios al productor para combustible y productos relacionados
$P$ = deflactor de precios del sector de negocios privado
$t$ = tiempo

Los números en paréntesis son los estadísticos $t$.

*a*) ¿Apoyan los resultados la hipótesis del autor?

*b*) Entre 1972 y 1977, el precio relativo de la energía ($P_e/P$) aumentó 60%. A partir de la regresión estimada, ¿cuál es la pérdida en productividad?

*c*) Después de permitir los cambios en ($h/k$) y ($P_e/P$), ¿cuál fue la tendencia de la tasa de crecimiento de la productividad durante el periodo muestral?

*d*) ¿Cómo interpreta el valor del coeficiente de 0.7135?

*e*) ¿El hecho de que cada coeficiente de pendiente parcial estimado sea estadísticamente significativo en el nivel individual (¿por qué?) significa que podemos rechazar la hipótesis de que $R^2 = 0$? ¿Por qué?

8.26. *La demanda de cable*. La tabla 8.10 presenta los datos de un fabricante de cable telefónico para pronosticar las ventas a uno de sus principales clientes durante el periodo 1968-1983.†

Las variables en la tabla se definen de la siguiente forma:

$Y$ = ventas anuales en millones de pies de cables pareados (MPC)
$X_2$ = Producto Interno Bruto (PIB), $, miles de millones
$X_3$ = construcción de nuevas viviendas, miles de unidades
$X_4$ = tasa de desempleo, %
$X_5$ = tasa preferencial rezagada 6 meses
$X_6$ = ganancias de línea para el cliente, %

* Consulte "Energy Prices and Capital Formation: 1972-1977", *Review*, Banco de la Reserva Federal de St. Louis, vol. 61, núm. 5, mayo de 1979, p. 4.

† El autor agradece a Daniel J. Reardon por recopilar y procesar los datos.

**TABLA 8.10** **Variables de regresión**

| Año | $X_2$, PIB | $X_3$, construcción de nuevas viviendas | $X_4$, desempleo % | $X_5$, tasa preferencial, rezago 6 meses | $X_6$, ganancias línea cliente, % | $Y$, ventas anuales (MPF) |
|---|---|---|---|---|---|---|
| 1968 | 1 051.8 | 1 503.6 | 3.6 | 5.8 | 5.9 | 5 873 |
| 1969 | 1 078.8 | 1 486.7 | 3.5 | 6.7 | 4.5 | 7 852 |
| 1970 | 1 075.3 | 1 434.8 | 5.0 | 8.4 | 4.2 | 8 189 |
| 1971 | 1 107.5 | 2 035.6 | 6.0 | 6.2 | 4.2 | 7 497 |
| 1972 | 1 171.1 | 2 360.8 | 5.6 | 5.4 | 4.9 | 8 534 |
| 1973 | 1 235.0 | 2 043.9 | 4.9 | 5.9 | 5.0 | 8 688 |
| 1974 | 1 217.8 | 1 331.9 | 5.6 | 9.4 | 4.1 | 7 270 |
| 1975 | 1 202.3 | 1 160.0 | 8.5 | 9.4 | 3.4 | 5 020 |
| 1976 | 1 271.0 | 1 535.0 | 7.7 | 7.2 | 4.2 | 6 035 |
| 1977 | 1 332.7 | 1 961.8 | 7.0 | 6.6 | 4.5 | 7 425 |
| 1978 | 1 399.2 | 2 009.3 | 6.0 | 7.6 | 3.9 | 9 400 |
| 1979 | 1 431.6 | 1 721.9 | 6.0 | 10.6 | 4.4 | 9 350 |
| 1980 | 1 480.7 | 1 298.0 | 7.2 | 14.9 | 3.9 | 6 540 |
| 1981 | 1 510.3 | 1 100.0 | 7.6 | 16.6 | 3.1 | 7 675 |
| 1982 | 1 492.2 | 1 039.0 | 9.2 | 17.5 | 0.6 | 7 419 |
| 1983 | 1 535.4 | 1 200.0 | 8.8 | 16.0 | 1.5 | 7 923 |

Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t$$

*a*) Estime la regresión anterior.

*b*) ¿Cuáles son los signos esperados para los coeficientes de este modelo?

*c*) ¿Corresponden los resultados empíricos a las expectativas *a priori*?

*d*) ¿Son los coeficientes de regresión parcial estimados estadísticamente significativos considerados en forma individual en el nivel de 5% de significancia?

*e*) Suponga que efectúa la regresión de $Y$ sobre $X_2$, $X_3$ y $X_4$ solamente y luego decide agregar las variables $X_5$ y $X_6$. ¿Cómo averiguará si se justifica agregar las variables $X_5$ y $X_6$? ¿Qué prueba utiliza? Muestre los cálculos necesarios.

8.27. Marc Nerlove estimó la siguiente función de costo para la generación de electricidad:*

$$Y = AX^{\beta} P^{\alpha_1} P^{\alpha_2} P^{\alpha_3} u \qquad (1)$$

donde $Y$ = costo total de producción
$X$ = producción en horas kilowatt
$P_1$ = precio del insumo trabajo
$P_2$ = precio del insumo capital
$P_3$ = precio del combustible
$u$ = término de perturbación

* Marc Nerlove, "Returns to Scale in Electric Supply", en Carl Christ (ed.), *Measurement in Economics*, Stanford University Press, Palo Alto, California, 1963. La notación cambió.

En teoría, se espera que la suma de las elasticidades del precio sea igual a la unidad, es decir, $(\alpha_1 + \alpha_2 + \alpha_3) = 1$. Pero al imponer esta restricción, la función de costos anterior se escribe como

$$(Y/P_3) = AX^{\beta}(P_1/P_3)^{\alpha_1}(P_2/P_3)^{\alpha_2}u \tag{2}$$

En otras palabras, (1) es una función de costo no restringida y (2) es una función de costo restringida.

Con base en una muestra de 29 empresas de tamaño mediano y después de realizar la transformación logarítmica, Nerlove obtuvo los siguientes resultados de la regresión:

$$\begin{aligned} \widehat{\ln Y_i} &= -4.93 \quad + \ 0.94 \ln X_i + \ 0.31 \ln P_1 \\ \text{ee} &= \ (1.96) \qquad (0.11) \qquad (0.23) \\ &\quad -0.26 \ln P_2 + \ 0.44 \ln P_3 \\ &\quad\ (0.29) \qquad (0.07) \qquad\qquad \text{SCR} = 0.336 \end{aligned} \tag{3}$$

$$\begin{aligned} \widehat{\ln(Y/P_3)} &= -6.55 + 0.91 \ln X + \ 0.51 \ln(P_1/P_3) + \ 0.09 \ln(P_2/P_3) \\ \text{ee} &= \ (0.16) \ \ (0.11) \qquad (0.19) \qquad\qquad (0.16) \qquad \text{SCR} = 0.364 \end{aligned} \tag{4}$$

*a*) Interprete las ecuaciones (3) y (4).

*b*) ¿Cómo averiguaría si la restricción $(\alpha_1 + \alpha_2 + \alpha_3) = 1$ es válida? Muestre sus cálculos.

8.28. *Estimación del modelo de asignación de precios de activos de capital (CAPM)*. En la sección 6.1 consideramos brevemente el conocido modelo de asignación de precios de activos de capital de la teoría moderna de portafolios. En el análisis empírico, el CAPM se estima en dos etapas.

**Etapa I (Regresión de serie de tiempo).** Para cada uno de los *N* títulos incluidos en la muestra efectuamos la siguiente regresión a través del tiempo:

$$R_{it} = \hat{\alpha}_i + \hat{\beta}_i R_{mt} + e_{it} \tag{1}$$

donde $R_{it}$ y $R_{mt}$ son las tasas de rendimiento del *i*-ésimo título y el portafolios del mercado (por ejemplo, el S&P 500) en el año *t*; $\beta_i$, como ya vimos, es el coeficiente beta o coeficiente de volatilidad del mercado del *i*-ésimo título y $e_{it}$ son los residuos. En total hay *N* regresiones, una para cada título, y se producen, por consiguiente, *N* valores estimados para $\beta_i$.

**Etapa II (Regresión transversal).** En esta etapa efectuamos la siguiente regresión para los *N* títulos:

$$\bar{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2 \hat{\beta}_i + u_i \tag{2}$$

donde $\bar{R}_i$ es el promedio o tasa media de rendimiento para el título *i*, calculado sobre el periodo muestral cubierto por la etapa I, $\hat{\beta}_i$ es el coeficiente beta estimado de la regresión de la primera etapa y $u_i$ es el término residual.

Al comparar la regresión (2) de la segunda etapa con el CAPM, ecuación (6.1.2), escrita como

$$\text{ER}_i = r_f + \beta_i(\text{ER}_m - r_f) \tag{3}$$

donde $r_f$ es la tasa de rendimiento libre de riesgo, vemos que $\hat{\gamma}_1$ es una estimación de $r_f$ y es $\hat{\gamma}_2$ una estimación de $(\text{ER}_m - r_f)$, la prima del riesgo del mercado.

Así, en la prueba empírica de CAPM, $\bar{R}_i$ y $\hat{\beta}_i$ se utilizan como estimadores de $ER_i$ y $\beta_i$ respectivamente. Ahora, si se mantiene CAPM, estadísticamente,

$$\hat{\gamma}_1 = r_f$$

$$\hat{\gamma}_2 = R_m - r_f, \text{ el estimador de } (ER_m - r_f)$$

Considere ahora otro modelo:

$$\bar{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2 \hat{\beta}_i + \hat{\gamma}_3 s^2_{e_i} + u_i \tag{4}$$

donde $s^2_{e_i}$ es la varianza residual del $i$-ésimo título de la regresión de la primera etapa. Entonces, si CAPM es válido, $\hat{\gamma}_3$ no debe ser significativamente diferente de cero.

Para probar el CAPM, Levy efectuó las regresiones (2) y (4) sobre una muestra de 101 acciones durante el periodo 1948-1968 y obtuvo los siguientes resultados:*

$$\begin{aligned} \hat{\bar{R}}_i &= \underset{(0.009)}{0.109} + \underset{(0.008)}{0.037}\beta_i \\ t &= (12.0) \qquad (5.1) \qquad R^2 = 0.21 \end{aligned} \tag{2$'$}$$

$$\begin{aligned} \hat{\bar{R}}_i &= \underset{(0.008)}{0.106} + \underset{(0.007)}{0.0024}\hat{\beta}_i + \underset{(0.038)}{0.201}s^2_{ei} \\ t &= (13.2) \qquad (3.3) \qquad (5.3) \qquad R^2 = 0.39 \end{aligned} \tag{4$'$}$$

*a*) ¿Apoyan estos resultados el CAPM?

*b*) ¿Se justifica agregar la variable $s^2_{e_i}$ al modelo? ¿Cómo sabe?

*c*) Si el CAPM se mantiene, $\hat{\gamma}_1$ en (2)′ debe aproximar el valor promedio de la tasa libre de riesgo $r_f$. El valor estimado es 10.9%. ¿Parece una estimación razonable de la tasa de rendimiento libre de riesgo durante el periodo de observación, 1948-1968? (Se puede considerar la tasa de rendimiento de los bonos del Tesoro o de un activo libre de riesgo relativamente parecido.)

*d*) Si el CAPM se mantiene, la prima de riesgo del mercado $(\bar{R}_m - r_f)$ de (2)′ es cerca de 3.7%. Si se supone que $r_f$ es 10.9%, esto implica que $\bar{R}_m$ para el periodo de la muestra fue aproximadamente 14.6%. ¿Parece una estimación razonable?

*e*) ¿Qué puede decir sobre el CAPM en general?

8.29. Consulte el ejercicio 7.21c. Ahora que ya cuenta con las herramientas necesarias, ¿cuál(es) prueba(s) utilizaría(n) para elegir entre los dos modelos? Muestre los cálculos necesarios. Observe que las variables dependientes en los dos modelos son distintas.

8.30. Consulte el ejemplo 8.3. Utilice la prueba $t$, como se muestra en (8.6.4), para averiguar si hubo rendimientos constantes a escala en la economía mexicana para el periodo de estudio.

8.31. Vuelva al ejemplo de la mortalidad infantil que estudiamos en diversas ocasiones. La regresión (7.6.2) se llevó a cabo para la mortalidad infantil (MI) sobre el PIB *per cápita* (PIBPC) y la tasa de alfabetización de las mujeres (TAM). Ahora extienda este modelo

* H. Levy, "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio", *American Economic Review*, vol. 68, núm. 4, septiembre de 1978, pp. 643-658.

para incluir la tasa de fecundidad total (TFT). Los datos de todas estas variables ya se dieron en la tabla 6.4. Reproducimos la regresión (7.6.2) y presentamos los resultados del modelo de regresión extendido:

1. $\widehat{MI}_i = 263.6416 - 0.0056\ \text{PIBPC}_i - 2.2316\ \text{TAM}_i$

$\text{ee} = (11.5932) \quad (0.0019) \quad (0.2099) \quad R^2 = 0.7077$

2. $\widehat{MI}_i = 168.3067 - 0.0055\ \text{PIBPC}_i - 1.7680\ \text{TAM}_i + 12.8686\ \text{TFT}_i$

$\text{ee} = (32.8916) \quad (0.0018) \quad (0.2480) \quad (?)$

$R^2 = 0.7474$

*a*) ¿Cómo interpretaría el coeficiente de TFT? *A priori*, ¿esperaría una relación positiva o negativa entre MI y TFT? Justifique su respuesta.

*b*) ¿Los valores de los coeficientes de PIBPC y TFT cambiaron entre ambas ecuaciones? Si así fue, ¿cuál(es) sería(n) la(s) razón(razones) de ese cambio? ¿La diferencia que se observa es estadísticamente significativa? ¿Qué prueba utilizaría y por qué?

*c*) ¿Cómo elegiría entre los modelos 1 y 2? ¿Qué prueba estadística emplearía para responder esta pregunta? Muestre los cálculos necesarios.

*d*) No se ha proporcionado el error estándar del coeficiente de TFT. ¿Puede calcularlo? (*Sugerencia:* Recuerde la relación entre las distribuciones $t$ y $F$.)

8.32. Consulte el ejercicio 1.7, en el que se dieron datos sobre los impactos publicitarios que se recuerdan y los gastos publicitarios para una muestra de 21 empresas. En el ejercicio 5.11 se pidió graficar esos datos y elaborar un modelo apropiado sobre la relación entre dichas variables. Con $Y$ como los impactos publicitarios recordados y $X$ el gasto publicitario, se obtuvieron las siguientes regresiones:

Modelo I: $\hat{Y}_i = 22.163 + 0.3631X_i$

$\text{ee} = (7.089) \quad (0.0971) \quad r^2 = 0.424$

Modelo II: $\hat{Y}_i = 7.059 + 1.0847X_i - 0.0040X_i^2$

$\text{ee} = (9.986) \quad (0.3699) \quad (0.0019) \quad R^2 = 0.53$

*a*) Interprete ambos modelos

*b*) ¿Cuál es el mejor modelo y por qué?

*c*) ¿Qué prueba(s) estadística(s) utilizaría para elegir entre los dos modelos?

*d*) ¿Hay "rendimientos decrecientes" del gasto publicitario?, es decir, ¿después de un determinado nivel de gasto publicitario (el nivel de saturación) ya no conviene gastar en publicidad? ¿Puede descubrir cuál sería ese nivel de gasto? Muestre los cálculos necesarios.

8.33. En la regresión (7.9.4) presentamos los resultados de la función de producción Cobb-Douglas ajustados al sector manufacturero de los 50 estados de Estados Unidos y Washington, D.C., para 2005. Con base en esa regresión determine si hay rendimientos constantes a escala en ese sector, mediante:

*a*) La prueba $t$ dada en (8.6.4). La covarianza entre los dos estimadores de pendiente es −0.03843.

*b*) La prueba $F$ dada en (8.6.9).

*c*) ¿Existe alguna diferencia en los resultados de estas dos pruebas? ¿Cuál es su conclusión respecto de los rendimientos a escala en el sector manufacturero de los 50 estados y Washington, D.C., en el periodo de muestra?

8.34. Reconsidere la regresión ahorro-ingreso de la sección 8.7. Suponga que dividimos la muestra en dos periodos: 1970-1982 y 1983-1995. Mediante la prueba de Chow determine si existe un cambio estructural en la regresión ahorro-ingreso para los dos periodos. Al comparar los resultados con los obtenidos en la sección 8.7, ¿qué conclusiones generales deduce respecto de la sensibilidad de la prueba de Chow a la elección del punto crítico que divide la muestra en dos (o más) periodos?

8.35. Consulte el ejercicio 7.24 y los datos de la tabla 7.12 relativos a cuatro variables económicas en Estados Unidos durante el periodo 1947-2000.

*a*) Con base en la regresión del gasto de consumo sobre el ingreso real, la riqueza real y la tasa de interés real, averigüe qué coeficientes de regresión son estadísticamente significativos, en lo individual, en el nivel de significancia de 5%. ¿Los signos de los coeficientes estimados concuerdan con la teoría económica?

*b*) Con base en los resultados de *a*), ¿cómo estimaría las elasticidades del ingreso, riqueza y tasa de interés? ¿Qué información adicional, si acaso, necesita para calcular las elasticidades?

*c*) ¿Cómo probaría la hipótesis de que las elasticidades del ingreso y la riqueza son iguales? Muestre los cálculos necesarios.

*d*) Suponga que en lugar de la función lineal de consumo que estimó en *a*), hace la regresión del logaritmo del gasto de consumo sobre los logaritmos del ingreso y de la riqueza y la tasa de interés. Obtenga los resultados de la regresión. ¿Cómo interpretaría estos resultados?

*e*) ¿Cuáles son las elasticidades del ingreso y la riqueza estimadas en *d*)? ¿Cómo interpreta el coeficiente de la tasa de interés estimado en *d*)?

*f*) En la regresión realizada en *d*), ¿pudo haber usado el logaritmo de la tasa de interés en lugar de la tasa de interés? ¿Por qué?

*g*) ¿Cómo compara las elasticidades estimadas en *b*) y en *d*)?

*h*) Entre los modelos de regresión estimados en *a*) y *d*), ¿cuál preferiría? ¿Por qué?

*i*) Suponga que en lugar de estimar el modelo dado en *d*), sólo hace la regresión del logaritmo del gasto de consumo sobre el logaritmo del ingreso. ¿Cómo decidiría si vale la pena agregar el logaritmo de la riqueza al modelo? ¿Y cómo decidiría si vale la pena agregar tanto el logaritmo de la riqueza como la tasa de interés al modelo? Muestre los cálculos necesarios.

8.36. Consulte la sección 8.8 y los datos de la tabla 8.9 relativos al ingreso personal disponible y el ahorro personal durante el periodo 1970-1995. En esa sección se introdujo la prueba de Chow para ver si ocurría un cambio estructural dentro de los datos entre dos periodos. La tabla 8.11 incluye datos actualizados con los valores de 1970-2005. Según el National Bureau of Economic Research, el ciclo de contracción del sector empresarial estadounidense más reciente terminó a finales de 2001. Divida los datos en tres secciones: 1) 1970-1981, 2) 1982-2001 y 3) 2002-2005.

*a*) Estime tanto el modelo correspondiente al conjunto completo de datos (años 1970-2005) como el correspondiente a la tercera sección (a partir de 2002). Use la prueba de Chow para determinar si existe una ruptura significativa entre el tercer periodo y el conjunto completo de datos.

*b*) Con los nuevos datos de la tabla 8.11 determine si todavía existe diferencia significativa entre el primer grupo de años (1970-1981) y el conjunto completo de datos ahora que dispone de más observaciones.

*c*) Ejecute la prueba de Chow con el periodo intermedio (1982-2001) en relación con el conjunto completo de datos para ver si los datos de este periodo muestran diferencias significativas respecto del resto de los datos.

**TABLA 8.11**
**Ahorro e ingreso personal disponible (miles de millones de dólares), Estados Unidos, 1970-2005 (miles de millones de dólares, salvo que se indique otra cosa; datos trimestrales con tasas anuales ajustadas por estacionalidad)**

Fuente: Department of Commerce, Bureau of Economic Analysis.

| Año | Ahorro | Ingreso |
|---|---|---|
| 1970 | 69.5 | 735.7 |
| 1971 | 80.6 | 801.8 |
| 1972 | 77.2 | 869.1 |
| 1973 | 102.7 | 978.3 |
| 1974 | 113.6 | 1 071.6 |
| 1975 | 125.6 | 1 187.4 |
| 1976 | 122.3 | 1 302.5 |
| 1977 | 125.3 | 1 435.7 |
| 1978 | 142.5 | 1 608.3 |
| 1979 | 159.1 | 1 793.5 |
| 1980 | 201.4 | 2 009.0 |
| 1981 | 244.3 | 2 246.1 |
| 1982 | 270.8 | 2 421.2 |
| 1983 | 233.6 | 2 608.4 |
| 1984 | 314.8 | 2 912.0 |
| 1985 | 280.0 | 3 109.3 |
| 1986 | 268.4 | 3 285.1 |
| 1987 | 241.4 | 3 458.3 |
| 1988 | 272.9 | 3 748.7 |
| 1989 | 287.1 | 4 021.7 |
| 1990 | 299.4 | 4 285.8 |
| 1991 | 324.2 | 4 464.3 |
| 1992 | 366.0 | 4 751.4 |
| 1993 | 284.0 | 4 911.9 |
| 1994 | 249.5 | 5 151.8 |
| 1995 | 250.9 | 5 408.2 |
| 1996 | 228.4 | 5 688.5 |
| 1997 | 218.3 | 5 988.8 |
| 1998 | 276.8 | 6 395.9 |
| 1999 | 158.6 | 6 695.0 |
| 2000 | 168.5 | 7 194.0 |
| 2001 | 132.3 | 7 486.8 |
| 2002 | 184.7 | 7 830.1 |
| 2003 | 174.9 | 8 162.5 |
| 2004 | 174.3 | 8 681.6 |
| 2005 | 34.8 | 9 036.1 |

## *Apéndice 8A2

# Prueba de la razón de verosimilitud (RV)

La **prueba de la RV** se basa en el principio de máxima verosimilitud (MV) estudiado en el apéndice 4A, en el cual se muestra la forma de obtener los estimadores de MV del modelo de regresión con dos variables. Ese principio puede extenderse directamente al modelo de regresión múltiple. Conforme al supuesto de que las perturbaciones $u_i$ están normalmente distribuidas, se muestra que para el modelo de regresión con dos variables los estimadores de MCO y MV de los coeficientes de regresión son idénticos, pero las varianzas

* Opcional.

del error estimado son diferentes. El estimador de MCO de $\sigma^2$ es $\sum \hat{u}_i^2/(n-2)$, pero el estimador MV es $\sum \hat{u}_i^2/n$, el primero insesgado y el último sesgado, aunque en muestras grandes el sesgo tiende a desaparecer.

Sucede lo mismo en la regresión múltiple. Para ilustrar lo anterior considere el modelo de regresión de tres variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{1}$$

Correspondiente a la ecuación (5) del apéndice 4A, el logaritmo de la función de verosimilitud para el modelo (1) se expresa así:

$$\ln \text{FV} = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \tag{2}$$

Como se muestra en el apéndice 4A, al diferenciar esta función respecto de $\beta_1$, $\beta_2$, $\beta_3$ y $\sigma^2$, igualar la expresión resultante a cero y resolver, se obtienen los estimadores de MV de tales estimadores. Los estimadores de MV de $\beta_1$, $\beta_2$ y $\beta_3$ serán idénticos a los estimadores de MCO, que ya se dieron en las ecuaciones (7.4.6) a (7.4.8), pero la varianza del error será diferente, pues la suma de cuadrados residual (SCR) estará dividida por $n$, en vez de $(n-3)$, como en el caso de MCO.

Ahora, supongamos que la hipótesis nula $H_0$ es que $\beta_3$, el coeficiente de $X_3$, es cero. En este caso, el log FV dado en (2) se convierte en

$$\ln \text{ FV} = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ \ln(2\pi) - \frac{1}{2\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_{2i})^2 \tag{3}$$

La ecuación (3) se conoce como el **logaritmo de la función de verosimilitud restringida (LFVR)** por estimarse con la restricción de que *a priori* $\beta_3$ es cero, mientras que la ecuación (2) se conoce como el logaritmo de la **función de verosimilitud no restringida (LFVNR)**, porque no se impusieron restricciones *a priori* sobre los parámetros. Para probar la validez de la restricción *a priori* de que $\beta_3$ es cero, la prueba de la RV produce el siguiente estadístico de prueba:

$$\lambda = 2(\text{LFVNR} - \text{LFVR}) \tag{4}^*$$

donde LFVNR y LFVR son el logaritmo de la función de verosimilitud no restringida [ecuación (2)] y el logaritmo de la función de verosimilitud restringida [ecuación (3)], respectivamente. Si el tamaño de la muestra es grande, puede demostrarse que el estadístico de prueba λ dado en (4) sigue una distribución ji cuadrada ($\chi^2$) con un número de gl igual al número de restricciones impuestas según la hipótesis nula, 1 en el presente caso.

La idea básica de la prueba de la RV es simple: si la(s) restricción(restricciones) *a priori* e (son) válida(s), los (log) FV restringida y no restringida no deben ser diferentes, en cuyo caso λ en (4) será cero. Pero si ése no es el caso, las dos FV divergirán. Como cuando la muestra es grande λ sigue una distribución ji cuadrada, es posible averiguar si la divergencia es estadísticamente significativa, por ejemplo, en un nivel de significancia de 1 o 5%. O de lo contrario, podemos encontrar el valor $p$ de la λ estimada.

Ilustremos la prueba de la RV con el ejemplo de mortalidad infantil. Si se hace la regresión de la mortalidad infantil (MI) sobre el PIB *per cápita* (PIBPC) y la tasa de alfabetización de las mujeres (TAM), como en (8.1.4), obtenemos una LFVNR de −328.1012, pero si sólo hacemos la regresión de MI sobre PIBPC, obtenemos una LFVR de −361.6396. En valor absoluto (es decir, sin considerar los signos), el primero es menor que el segundo, lo cual tiene sentido, pues tenemos una variable adicional en el primer modelo.

La cuestión ahora es si vale la pena añadir la variable TAM. Si no es así, los logaritmos de FV restringida y FV no restringida no diferirán mucho, pero, en caso contrario, los logaritmos de las FV serán muy distintos. Para apreciar si esta diferencia es estadísticamente significativa, utilizaremos la prueba de la RV dada en (4), lo cual origina:

$$\lambda = 2[-328.1012 - (-361.6396)] = 67.0768$$

---

* Esta expresión también se expresa como −2(LFVR − LFVNR) o −2 ln(FVR/FVNR).

Asintóticamente, su distribución es igual a la ji cuadrada con 1 gl (porque sólo se tiene una restricción impuesta cuando se omite la variable TAM del módulo completo). El valor $p$ de obtener tal valor ji cuadrada para 1 gl es casi cero, lo cual lleva a la conclusión de que la variable TAM *no* debe excluirse del modelo. En otras palabras, la regresión restringida en el presente ejemplo no es válida.

Si SCRR y SCRNR representan las sumas de cuadrados residuales restringida y no restringida, la ecuación (4) también se expresa como:

$$-2 \ln \lambda = n(\ln \text{SCRR} - \ln \text{SCRNR}) \tag{5}$$

que está distribuida como $\chi^2$ con $r$ grados de libertad, donde $r$ es el número de restricciones impuestas al modelo (es decir, el número de coeficientes $r$ omitidos del modelo original).

Aunque no es posible estudiar en detalle las pruebas de Wald y MV, estas pruebas pueden aplicarse como sigue:

$$\text{Estadístico de Wald (W)} = \frac{(n-k)(\text{SCRR} - \text{SCRNR})}{\text{SCRNR}} \sim \chi^2_r \tag{6}$$

$$\text{Estadístico del multiplicador de Lagrange (ML)} = \frac{(n-k+r)(\text{SCRR} - \text{SCRNR})}{\text{SCRR}} \sim \chi^2_r \tag{7}$$

donde $k$ es el número de regresoras en el modelo sin restricciones y $r$ es el número de restricciones.

Como puede observar en las ecuaciones anteriores, las tres pruebas son asintóticamente (es decir, en muestras grandes) equivalentes: producen respuestas similares. Sin embargo, las respuestas pueden diferir en muestras pequeñas. Existe una relación interesante entre estos estadísticos en que se puede demostrar que:

$$\text{W} \geq \text{RV} \geq \text{ML}$$

Por tanto, en muestras pequeñas, se puede rechazar una hipótesis con el estadístico de Wald, pero *no* con el estadístico del ML.*

Como se señaló en el texto, para casi todos los propósitos de este estudio bastan las pruebas $t$ y $F$. Pero las tres pruebas analizadas antes tienen aplicabilidad general en el sentido de probar hipótesis no lineales en modelos lineales, o probar restricciones sobre matrices de varianza-covarianza. También pueden aplicarse en situaciones donde no se sostiene el supuesto de que los errores están distribuidos normalmente.

Debido a la complejidad matemática de las pruebas de Wald y ML, no profundizaremos más aquí. Pero, como se anotó, asintóticamente las pruebas RV, Wald y ML producen respuestas idénticas; la elección entre una u otra prueba depende de la conveniencia computacional.

* Para una explicación, véase G.S. Maddala, *Introduction to Econometrics*, 3a. ed., John Wiley & Sons, Nueva York, 2001, p. 177.

# Capítulo 9

# Modelos de regresión con variables dicótomas

En el capítulo 1 analizamos brevemente los cuatro tipos de variables que por lo general se encuentran en el análisis empírico: **escala de razón**, **escala de intervalo**, **escala ordinal** y **escala nominal**. Los tipos de variables de los capítulos anteriores fueron en esencia en *escala de razón*. Pero esto no debe dar la impresión de que los modelos de regresión sólo tratan con variables en escala de razón. Los modelos de regresión también trabajan con los demás tipos de variables que acabamos de mencionar. En este capítulo consideraremos modelos que tal vez no sólo tengan variables en escala de razón, sino también variables en **escala nominal**. Estas variables también se conocen como **variables indicadoras**, **variables categóricas**, **variables cualitativas** o **variables dicótomas**.[1]

## 9.1 Naturaleza de las variables dicótomas

En el análisis de regresión, la variable dependiente o regresada a menudo acusa influencia no sólo de variables en escala de razón (por ejemplo: ingreso, producción, precios, costos y estatura), sino también de variables cualitativas por naturaleza, o de escala nominal (como sexo, raza, color, religión, nacionalidad, región geográfica, cambios políticos y afiliación partidista). Por ejemplo, con los demás factores constantes, se ha visto que las trabajadoras ganan menos que sus pares masculinos, y que las personas de color ganan menos que las blancas.[2] Este patrón puede resultar de la discriminación sexual o racial, pero cualquiera que sea la razón, las variables cualitativas, como sexo y raza, sí influyen en la variable dependiente y es claro que deben incluirse en las explicativas, o regresoras.

Como tales variables suelen indicar la presencia o ausencia de una "cualidad" o atributo, como femenino o masculino, negro o blanco, católico o no católico, demócrata o republicano, son variables en *escala nominal* esencialmente. Una manera de "cuantificar" tales atributos es mediante variables artificiales que toman los valores 0 o 1, donde 1 indica la presencia (o posesión) de ese atributo y 0 su ausencia. Por ejemplo, 1 puede indicar que una persona es de sexo femenino y 0 que es de sexo masculino; o 1 puede indicar que una persona se graduó en la universidad y 0 que no lo ha hecho, y así en cada caso. Las variables que adquieren tales valores 0 y 1 se llaman

[1] Analizaremos las variables en escala ordinal en el capítulo 15.

[2] Hay una revisión de la demostración de este tema en Bruce E. Kaufman y Julie L. Hotchkiss, *The Economics of Labor Market*, 5a. ed., Dryden Press, Nueva York, 2000.

**variables dicótomas**.[3] *Tales variables son, por tanto, en esencia, un recurso para clasificar datos en categorías mutuamente excluyentes, como masculino o femenino.*

Las variables dicótomas pueden utilizarse en los modelos de regresión en forma tan fácil como las variables cuantitativas. De hecho, un modelo de regresión puede contener variables explicativas exclusivamente dicótomas o cualitativas, por naturaleza. Tales modelos se denominan **modelos de análisis de varianza (ANOVA)**.[4]

## 9.2 Modelos ANOVA

Para ilustrar los modelos ANOVA, considere el siguiente ejemplo.

**EJEMPLO 9.1**
*Salarios de maestros de escuelas públicas por región geográfica*

La tabla 9.1 proporciona datos sobre salarios (en dólares) de los maestros de escuelas públicas en 50 estados y el Distrito de Columbia para los años 2005-2006. Las 51 áreas se clasifican en tres regiones geográficas: 1) Noreste y Norte-centro (21 estados en total); 2) Sur (17 estados en total), y 3) Oeste (13 estados en total). Por el momento, no se preocupe por el formato de la tabla ni de los demás datos suministrados.

Suponga que deseamos averiguar si el salario promedio anual (SPA) de los maestros de escuelas públicas difiere en las tres áreas geográficas de Estados Unidos. Si tomamos el promedio aritmético simple de los salarios promedio de los maestros de las tres regiones, obtenemos los siguientes promedios para las tres regiones: $49 538.71 (Noreste y Norte-centro), $46 293.59 (Sur) y $48 104.62 (Oeste). Esos números difieren entre sí, pero, ¿son estadísticamente distintos? Existen varias técnicas estadísticas para comparar dos o más valores medios, lo cual por lo general se conoce como **análisis de varianza**.[5] Pero se logra lo mismo con el análisis de regresión.

Para ver lo anterior, considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_{3i} D_{3i} + u_i \tag{9.2.1}$$

donde $Y_i$ = salario (promedio) de los maestros de escuelas públicas en el estado $i$
$D_{2i}$ = 1 si el estado se encuentra en el Noreste o Norte-centro
= 0 para otra región del país
$D_{3i}$ = 1 si el estado es del Sur
= 0 para otra región del país

Observe que (9.2.1) es como cualquier modelo de regresión múltiple que se haya estudiado antes, excepto que en vez de regresoras cuantitativas, se tienen sólo variables cualitativas o dicótomas, las cuales toman el valor de 1 si la observación pertenece a una categoría particular, y

[3] No es absolutamente esencial que las variables dicótomas adquieran los valores 0 y 1. El par (0,1) puede transformarse en cualquier otro par mediante una función lineal tal que $Z = a + bD$ ($b \neq 0$), donde $a$ y $b$ son constantes y donde $D = 1$ o 0. Cuando $D = 1$ se tiene $Z = a + b$, y cuando $D = 0$, se tiene $Z = a$. Así, el par (0,1) se convierte en ($a$, $a + b$). Por ejemplo, si $a = 1$ y $b = 2$, las variables dicótomas serán (1, 3). *Esta expresión muestra que las variables cualitativas o dicótomas no tienen una escala natural de medición.* Esto se debe a que se describen como variables en escala nominal.

[4] Los modelos ANOVA se utilizan para evaluar la significancia estadística de la relación entre una regresada cuantitativa y regresoras cualitativas o dicótomas. A menudo se emplean para comparar las diferencias entre los valores medios de dos o más grupos o categorías y, por tanto, son más generales que la prueba $t$, con la cual se comparan las medias de sólo dos grupos o categorías.

[5] Para un tratamiento de las aplicaciones, véase John Fox, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage Publications, 1997, cap. 8.

**TABLA 9.1 Salario promedio de maestros de escuelas públicas, por estado, 2005-2006**

| | Salario | Gasto | $D_2$ | $D_3$ | | Salario | Gasto | $D_2$ | $D_3$ |
|---|---|---|---|---|---|---|---|---|---|
| Connecticut | 60 822 | 12 436 | 1 | 0 | Georgia | 49 905 | 8 534 | 0 | 1 |
| Illinois | 58 246 | 9 275 | 1 | 0 | Kentucky | 43 646 | 8 300 | 0 | 1 |
| Indiana | 47 831 | 8 935 | 1 | 0 | Louisiana | 42 816 | 8 519 | 0 | 1 |
| Iowa | 43 130 | 7 807 | 1 | 0 | Maryland | 56 927 | 9 771 | 0 | 1 |
| Kansas | 43 334 | 8 373 | 1 | 0 | Mississippi | 40 182 | 7 215 | 0 | 1 |
| Maine | 41 596 | 11 285 | 1 | 0 | North Carolina | 46 410 | 7 675 | 0 | 1 |
| Massachusetts | 58 624 | 12 596 | 1 | 0 | Oklahoma | 42 379 | 6 944 | 0 | 1 |
| Michigan | 54 895 | 9 880 | 1 | 0 | South Carolina | 44 133 | 8 377 | 0 | 1 |
| Minnesota | 49 634 | 9 675 | 1 | 0 | Tennessee | 43 816 | 6 979 | 0 | 1 |
| Missouri | 41 839 | 7 840 | 1 | 0 | Texas | 44 897 | 7 547 | 0 | 1 |
| Nebraska | 42 044 | 7 900 | 1 | 0 | Virginia | 44 727 | 9 275 | 0 | 1 |
| New Hampshire | 46 527 | 10 206 | 1 | 0 | West Virginia | 40 531 | 9 886 | 0 | 1 |
| Nueva Jersey | 59 920 | 13 781 | 1 | 0 | Alaska | 54 658 | 10 171 | 0 | 0 |
| Nueva York | 58 537 | 13 551 | 1 | 0 | Arizona | 45 941 | 5 585 | 0 | 0 |
| Dakota del Norte | 38 822 | 7 807 | 1 | 0 | California | 63 640 | 8 486 | 0 | 0 |
| Ohio | 51 937 | 10 034 | 1 | 0 | Colorado | 45 833 | 8 861 | 0 | 0 |
| Pennsylvania | 54 970 | 10 711 | 1 | 0 | Hawaii | 51 922 | 9 879 | 0 | 0 |
| Rhode Island | 55 956 | 11 089 | 1 | 0 | Idaho | 42 798 | 7 042 | 0 | 0 |
| Dakota del Sur | 35 378 | 7 911 | 1 | 0 | Montana | 41 225 | 8 361 | 0 | 0 |
| Vermont | 48 370 | 12 475 | 1 | 0 | Nevada | 45 342 | 6 755 | 0 | 0 |
| Wisconsin | 47 901 | 9 965 | 1 | 0 | New Mexico | 42 780 | 8 622 | 0 | 0 |
| Alabama | 43 389 | 7 706 | 0 | 1 | Oregon | 50 911 | 8 649 | 0 | 0 |
| Arkansas | 44 245 | 8 402 | 0 | 1 | Utah | 40 566 | 5 347 | 0 | 0 |
| Delaware | 54 680 | 12 036 | 0 | 1 | Washington D.C. | 47 882 | 7 958 | 0 | 0 |
| Distrito de Columbia | 59 000 | 15 508 | 0 | 1 | Wyoming | 50 692 | 11 596 | 0 | 0 |
| Florida | 45 308 | 7 762 | 0 | 1 | | | | | |

*Nota:* $D_2 = 1$ para estados del Noreste y Norte-centro; 0 para otra región.
$D_3 = 1$ para estados del Sur; 0 para otra región.

Fuente: National Education Association como se informó en 2007.

0 si no pertenece a esa categoría o grupo. *De aquí en adelante, designaremos todas las variables dicótomas con la letra D.* La tabla 9.1 muestra las variables dicótomas así definidas.

¿Qué expresa el modelo (9.2.1)? Si consideramos que el término de error satisface las suposiciones usuales de MCO, al calcular la esperanza de (9.2.1) en ambos lados, obtenemos:

Salario medio de los maestros de escuelas públicas en la región Noreste y Norte-centro:

$$E(Y_i \mid D_{2i} = 1, D_{3i} = 0) = \beta_1 + \beta_2 \quad \textbf{(9.2.2)}$$

Salario medio de los maestros de escuelas públicas en el Sur:

$$E(Y_i \mid D_{2i} = 0, D_{3i} = 1) = \beta_1 + \beta_3 \quad \textbf{(9.2.3)}$$

Quizá se pregunte cómo calcular el salario promedio de los maestros de escuelas públicas en el Oeste. Si sospecha que la respuesta es $\beta_1$, está en lo correcto, pues:

El salario medio de los maestros de escuelas públicas en el Oeste:

$$E(Y_i \mid D_{2i} = 0, D_{3i} = 0) = \beta_1 \quad \textbf{(9.2.4)}$$

(*continúa*)

**EJEMPLO 9.1**
(*continuación*)

En otras palabras, el salario medio de los maestros de escuelas públicas en el Oeste está dado por el intercepto, $\beta_1$, en la regresión múltiple (9.2.1); además, los coeficientes de la "pendiente" $\beta_2$ y $\beta_3$ indican la cantidad por la que los salarios promedio de los maestros del Noreste y Norte-centro, así como los del Sur, difieren respecto de los salarios medios de los profesores del Oeste. Pero, ¿cómo saber si estas diferencias son estadísticamente significativas? Antes de responder, veamos los resultados basados en la regresión (9.2.1). Con los datos de la tabla 9.1 obtenemos los siguientes resultados:

$$\begin{aligned}
\hat{Y}_i &= 48\,014.615 \quad + 1\,524.099D_{2i} - 1\,721.027D_{3i} \\
\text{ee} &= (1\,857.204) \quad (2\,363.139) \quad (2\,467.151) \\
t &= (25.853) \quad (0.645) \quad (-0.698) \\
&\quad (0.0000)^* \quad (0.5220)^* \quad (0.4888)^* \qquad R^2 = 0.0440
\end{aligned} \tag{9.2.5}$$

donde * indica los valores $p$.

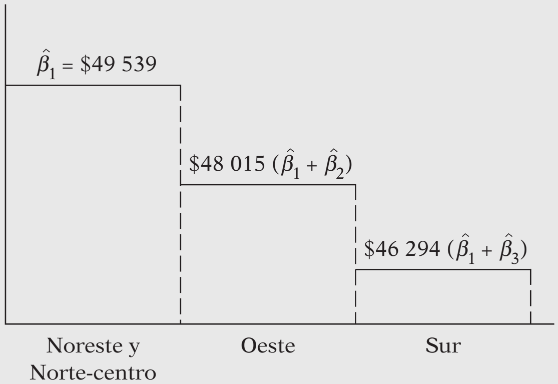
Como muestran los resultados de esta regresión, el salario medio de los profesores del Oeste es de casi \$48 015, el de los maestros del Noreste y del Norte-centro es mayor por cerca de \$1 524, y respecto de los del Sur, es menor por cerca de \$1 721. Los salarios medios reales en las últimas dos regiones se obtienen con facilidad si sumamos estos salarios diferenciales al salario medio de los maestros del Oeste, como se ve en las ecuaciones (9.2.3) y (9.2.4). Al hacer esto, tendremos que los salarios medios de las dos últimas regiones son cercanos a \$49 539 y \$46 294.

Pero, ¿cómo sabemos que estos salarios medios son estadísticamente diferentes del salario medio de los profesores del Oeste, que es la categoría con la que se comparan? Es muy fácil. Todo lo que hay que hacer es averiguar si cada coeficiente de "pendiente" en (9.2.5) es estadísticamente significativo. Como se observa en esta regresión, el coeficiente estimado de la pendiente para la región Noreste y Norte-centro no es estadísticamente significativo, pues su valor $p$ es 52%; tampoco el del Sur es estadísticamente significativo, pues el valor $p$ es más o menos de 49%. En consecuencia, la conclusión general es que, estadísticamente, los salarios medios de los profesores de escuelas públicas del Oeste, Noreste y Norte-centro, y Sur son casi iguales. La situación se ilustra en el diagrama de la figura 9.1.

Hay que tener cuidado al interpretar estas diferencias. Las variables dicótomas simplemente señalan las diferencias, si existen, pero no indican las razones por las que se presentan. Las diferencias en los niveles educativos, los índices del costo de vida, el sexo y la raza quizá ejerzan algún efecto sobre las diferencias observadas. Por tanto, a menos que se tomen en cuenta todas las demás variables que puedan afectar el salario de un maestro, no se podrán aclarar las causas de las diferencias.

Del análisis anterior, resulta claro que lo único que hay que hacer es observar si los coeficientes relacionados con las diferentes variables dicótomas son estadísticamente significativos en lo individual. Este ejemplo también muestra lo fácil que es incorporar regresoras cualitativas, o dicótomas, a los modelos de regresión.

**FIGURA 9.1**
Salario promedio (en dólares) de los maestros de escuelas públicas de tres regiones.



## Precaución con las variables dicótomas

Aunque es fácil añadirlas a los modelos de regresión, las variables dicótomas se deben utilizar con cuidado. En particular, considere los siguientes aspectos:

1. En el ejemplo 9.1, para diferenciar las tres regiones utilizamos sólo dos variables dicótomas, $D_2$ y $D_3$. ¿Por qué no empleamos tres variables dicótomas para distinguir las tres regiones? Suponga que hacemos precisamente eso y escribimos el modelo (9.2.1) como:

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.6}$$

donde $D_{1i}$ toma el valor de 1 para los estados del Oeste y 0 para los de otras regiones. Por tanto, ahora tenemos una variable dicótoma para cada una de las tres regiones geográficas. Con los datos de la tabla 9.1, si fuese a hacer la regresión de (9.2.6), la computadora "se negaría" (inténtelo).[6] ¿Por qué? La razón estriba en que cuando se definió (9.2.6), donde se tiene una variable dicótoma para cada categoría o grupo, así como un intercepto, se presenta un caso de **colinealidad perfecta**; es decir, existe una relación lineal exacta entre las variables. ¿Por qué? Consulte la tabla 9.1. Imagine que ahora añadimos la columna $D_1$, que toma el valor de 1 siempre que un estado sea del Oeste y 0 en cualquier otro caso. Ahora bien, si sumamos las tres columnas $D$ horizontalmente, obtendremos una columna con 51 números 1. Pero, como el valor del intercepto $\alpha$ es (implícitamente) 1 para cada observación, habrá una columna también con 51 números 1. En otras palabras, la suma de las tres columnas $D$ sólo reproducirá la columna del intercepto, lo cual provoca colinealidad perfecta. En este caso es imposible la estimación del modelo (9.2.6).

El mensajes es: **si una variable cualitativa tiene *m* categorías, sólo hay que agregar ($m - 1$) variables dicótomas**. En el ejemplo anterior, como la variable cualitativa "región" tiene tres categorías, se introducen sólo dos variables dicótomas. Si no se respeta esta regla se provocará lo que se conoce como **trampa de la variable dicótoma**; es decir, se tendrá una situación de perfecta colinealidad o perfecta multicolinealidad, si hay más de una relación exacta entre las variables. Esta regla también vale si se tiene más de una variable cualitativa en el modelo, sobre lo cual veremos un ejemplo más adelante. Así, se tiene que enunciar de nuevo la regla anterior como: **para cada regresora cualitativa, el número de variables dicótomas introducidas debe ser una menos que las categorías de esa variable**. Entonces, si en el ejemplo 9.1 hubiésemos contado con información sobre el género de los profesores, habríamos utilizado una variable dicótoma adicional (pero no dos) que tomara el valor de 1 para mujer y de 0 para hombre, o viceversa.

2. La categoría a la cual no se asigna variable dicótoma se conoce como **categoría base**, **de comparación**, **de control**, **de referencia** u **omitida**. Además, todas las comparaciones se hacen respecto de la categoría de comparación.

3. El valor del intercepto ($\beta_1$) representa el *valor medio* de la categoría de comparación. En el ejemplo 9.1, dicha categoría es la región Oeste. En consecuencia, para la regresión (9.2.5), el valor del intercepto, de alrededor de 48 015, representa el salario medio de los maestros para los estados del Oeste.

4. Los coeficientes asociados a las variables dicótomas en (9.2.1) se conocen como **coeficientes de intercepto diferencial**, debido a que indican la medida en que el valor de la categoría que recibe el valor de 1 difiere del coeficiente de intercepto correspondiente a la categoría de comparación. Por ejemplo, en (9.2.5), el valor aproximado de 1 524 señala que el salario promedio de los maestros de la región Noreste y Norte-centro es mayor por aproximadamente \$1 524 que el salario medio de casi \$48 015 perteneciente a la categoría de comparación, en este caso, el Oeste.

[6] En realidad, obtendría el mensaje de que la matriz es singular.

5. Si una variable cualitativa tiene más de una categoría, como en el ejemplo ilustrativo, la elección de la categoría de comparación se deja al criterio estricto del investigador. A veces dicha elección la determina el problema particular que se trabaja. Para el ejemplo ilustrativo se pudo elegir el Sur como categoría de comparación. En ese caso cambian los resultados de la regresión dados en (9.2.5), pues ahora las comparaciones se hacen respecto del Sur. Por supuesto, lo anterior no cambia la conclusión general del ejemplo (¿por qué?). En este caso, el valor del intercepto será cercano a $46 294, el salario medio de los maestros del Sur.

6. Advertimos sobre la trampa de la variable dicótoma. Existe una forma de eludirla al introducir tantas variables dicótomas como números de categorías tenga dicha variable, *siempre y cuando no se introduzca el intercepto en dicho modelo*. Así, si eliminamos el término del intercepto de (9.2.6) y consideramos el siguiente modelo

$$Y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.7}$$

no caeremos en la trampa de la variable dicótoma, pues no existe colinealidad perfecta. *Pero se debe asegurar de que, cuando haga esa regresión, utilice la opción "no intercepto" en el paquete de software.*

¿Cómo interpretamos la regresión (9.2.7)? Si toma la esperanza de (9.2.7), tendrá que:

$\beta_1 =$ salario medio de los maestros en el Oeste

$\beta_2 =$ salario medio de los maestros en el Noreste y Norte-centro

$\beta_3 =$ salario medio de los maestros en el Sur

En otras palabras, *con el intercepto eliminado y al permitir una variable dicótoma para cada categoría, obtenemos de manera directa los valores medios de las distintas categorías*. Los resultados de (9.2.7) para el ejemplo ilustrativo son los siguientes:

$$\begin{aligned}
\hat{Y}_i &= 48\,014.62D_{1i} \quad + 49\,538.71D_{2i} \quad + 46\,293.59D_{3i} \\
\text{ee} &= (1\,857.204) \qquad (1\,461.240) \qquad (1\,624.077) \\
t &= \quad (25.853)^* \qquad (33.902)^* \qquad (28.505)^* \\
& \qquad\qquad\qquad\qquad\qquad\qquad R^2 = 0.044
\end{aligned} \tag{9.2.8}$$

donde* indica que los valores $p$ de estas razones $t$ son muy pequeños.

Como se observa, los coeficientes de las variables dicótomas proporcionan de manera directa los valores medios (de los salarios) para las tres regiones: Oeste, Noreste y Norte-centro, y Sur.

7. ¿Cuál de los siguientes métodos es el mejor para introducir una variable dicótoma: 1) agregar una variable dicótoma para cada categoría y omitir el término del intercepto o 2) incluir el término del intercepto y añadir sólo $(m - 1)$ variables, donde $m$ es el número de categorías de la variable dicótoma? Como señala Kennedy:

> La mayoría de los investigadores piensan que es más conveniente la ecuación con intercepto porque les permite enfrentar de manera más sencilla las interrogantes que a menudo les interesan más; a saber, si la categorización genera una diferencia o no; y si lo hace, en qué medida. Si la categorización genera una diferencia, el grado de esta diferencia se mide directamente por las estimaciones de los coeficientes de las variables dicótomas. Probar si la categorización es o no es relevante se lleva a cabo mediante la prueba $t$ del coeficiente de una variable dicótoma, respecto de cero (o, de forma más general, una prueba $F$ sobre el conjunto apropiado de los coeficientes estimados de las variables dicótomas).[7]

[7] Peter Kennedy, *A Guide to Econometrics*, 4a. ed., MIT Press, Cambridge, Massachusetts, 1998, p. 223.

## 9.3 Modelos ANOVA con dos variables cualitativas

En la sección anterior estudiamos un modelo ANOVA con una variable cualitativa de tres categorías. En esta sección analizaremos otro modelo ANOVA, pero con dos variables cualitativas, además de destacar otros aspectos sobre este tipo de variables.

**EJEMPLO 9.2**
*Salarios por hora en relación con el estado civil y la región de residencia*

De una muestra de 528 personas tomada en mayo de 1985 se obtuvieron los siguientes resultados de regresión:[8]

$$\begin{aligned}
\hat{Y}_i &= \quad 8.8148 \quad + \; 1.0997D_{2i} - \quad 1.6729D_{3i} \\
\text{ee} &= \; (0.4015) \qquad (0.4642) \qquad\quad (0.4854) \\
t &= (21.9528) \qquad (2.3688) \qquad (-3.4462) \\
&\quad\; (0.0000)^* \qquad (0.0182)^* \qquad (0.0006)^* \\
&\qquad\qquad\qquad\qquad\qquad\qquad R^2 = 0.0322
\end{aligned} \tag{9.3.1}$$

donde $Y$ = salario por hora ($)
$D_2$ = estado civil; 1 si es casado, 0 en otro caso
$D_3$ = región de residencia; 1 si es del Sur, 0 en otro caso

y * denota los valores $p$.

En este ejemplo tenemos dos regresoras cualitativas, cada una con dos categorías. Por tanto, asignamos una variable dicótoma para cada categoría.

¿Cuál es la categoría de comparación en este caso? Obvio, son los no casados y con residencia fuera del Sur. En otras palabras, las personas no casadas y que no viven en el Sur forman la categoría omitida. Por consiguiente, todas las comparaciones se establecen respecto de este grupo. El salario medio por hora en esta categoría base es de casi $8.81. Respecto de ésta, el salario promedio por hora de los que están casados es mayor por casi $1.10, lo cual da un salario promedio real de $9.91 (= 8.81 + 1.10). En contraste, para los que viven en el Sur, su salario promedio por hora es menor por cerca de $1.67, lo cual da un salario promedio por hora de $7.14.

¿Los salarios promedio por hora anteriores son estadísticamente distintos en comparación con la categoría base? Sí lo son, pues todos los interceptos diferenciales son estadísticamente significativos: sus valores $p$ son muy bajos.

El punto que debe notarse en este ejemplo es el siguiente: *una vez que se va más allá de una variable cualitativa, se tiene que poner mucha atención a la categoría considerada como base, porque todas las comparaciones se llevan a cabo respecto de dicha categoría. Esto es especialmente importante cuando se tienen varias regresoras cualitativas y cada una de ellas presenta diversas categorías*. A estas alturas, el mecanismo de introducción de diversas variables cualitativas debe ser claro para el lector.

## 9.4 Regresión con una mezcla de regresoras cualitativas y cuantitativas: los modelos ANCOVA

Los modelos ANOVA del tipo que vimos en las dos secciones anteriores, aunque son comunes en áreas como sociología, psicología, educación e investigación de mercados, no son tan frecuentes en la economía. Por lo general, en la mayor parte de la investigación económica, un modelo de

[8] Los datos se tomaron del disco de datos de Arthur S. Goldberger, *Introductory Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1998. Ya se tomaron en cuenta esos datos en el capítulo 2.

regresión contiene diversas variables explicativas cuantitativas y otras cualitativas. Los modelos de regresión que muestran una mezcla de variables cuantitativas y cualitativas se llaman **modelos de análisis de covarianza (ANCOVA)**. Tales modelos representan una generalización de los modelos ANOVA en el sentido de que proporcionan un método para controlar estadísticamente los efectos de las regresoras cuantitativas (llamadas **covariantes** o **variables de control**) en un modelo con regresoras cuantitativas y cualitativas (o dicótomas). A continuación se ilustran los modelos ANCOVA.

**EJEMPLO 9.3**
*Salario de los maestros en relación con la región y el gasto en escuelas públicas por alumno*

Para motivar el análisis regresemos al ejemplo 9.1 afirmando que el salario promedio de los maestros de escuelas públicas no variará en las tres regiones si se toma en cuenta cualquier variable que no pueda estandarizarse en las tres regiones. Por ejemplo, piense en la variable *gasto en escuelas públicas erogado por las autoridades locales*, en vista de que la educación primaria es una cuestión sobre todo de carácter local y estatal. Para ver si éste es el caso, desarrollamos el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i \tag{9.4.1}$$

donde $Y_i$ = salario promedio anual de los maestros de escuelas públicas en el estado ($)
$X_i$ = gasto en escuelas públicas por alumno ($)
$D_{2i}$ = 1 si el estado es del Noreste o Norte-centro; 0 en otro caso
$D_{3i}$ = 1 si el estado es del Sur; 0 en otro caso

Los datos para $X$ se proporcionan en la tabla 9.1. Tenga presente que se considera al Oeste como la categoría de comparación. Asimismo, note que, además de las dos regresoras cualitativas, se tiene una variable cuantitativa, $X$, que en el contexto de los modelos ANCOVA se conoce como **covariante**, como dijimos antes.

De los datos mostrados en la tabla 9.1, los resultados del modelo (9.4.1) son los siguientes:

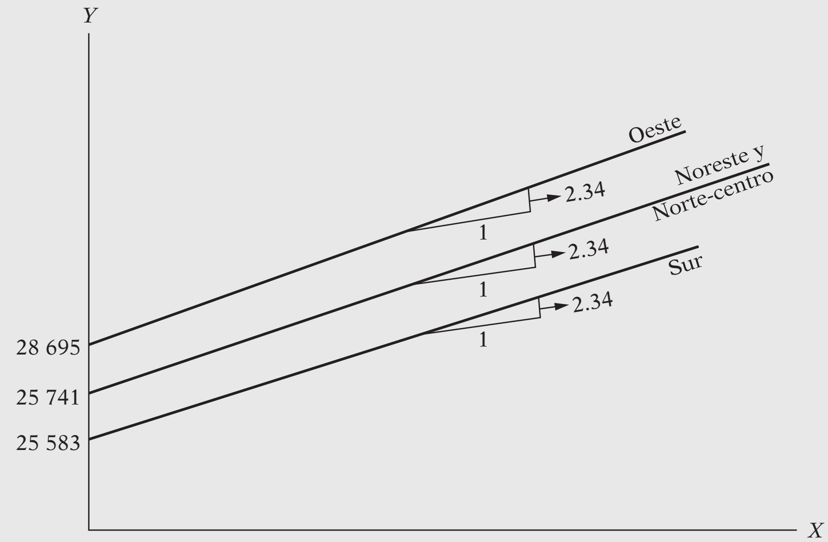
$$\begin{aligned}
\hat{Y}_i &= 28\,694.918 \quad - \quad 2\,954.127D_{2i} - 3\,112.194D_{3i} + \quad 2.3404X_i \\
\text{ee} &= (3\,262.521) \qquad (1\,862.576) \qquad (1\,819.873) \qquad (0.3592) \\
t &= \quad (8.795)^* \qquad (-1.586)^{**} \qquad (-1.710)^{**} \qquad (6.515)^* \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad R^2 = 0.4977
\end{aligned} \tag{9.4.2}$$

donde * indica valores $p$ menores que 5% y ** indica valores $p$ mayores que 5%.

Como los resultados indican, *ceteris paribus*: conforme el gasto público aumenta un dólar, en promedio, el salario de los maestros de escuela pública se incrementa más o menos $2.34. Si controlamos el gasto en educación, ahora se observa que el coeficiente de intercepto diferencial no es significativo para la región Noreste y Norte-centro ni para el Sur. Estos resultados difieren de los de (9.2.5). Pero no debe sorprender, pues en (9.2.5) no tuvimos en cuenta la covariante, que son las diferencias del gasto público en educación por alumno. La situación se ilustra de manera gráfica en la figura 9.2.

Note que, si bien se mostraron tres líneas de regresión para las tres regiones, estadísticamente las líneas de regresión son las mismas para las tres regiones. También observe que las tres líneas de regresión son paralelas. (¿Por qué?)

**FIGURA 9.2**
Salario de los maestros de escuelas públicas ($Y$) en relación con el gasto en educación por alumno ($X$).




## 9.5 La variable dicótoma alternativa a la prueba de Chow[9]

En la sección 8.7 analizamos la prueba de Chow para examinar la estabilidad estructural del modelo de regresión. El ejemplo ahí tuvo que ver con la relación entre ahorro e ingreso en Estados Unidos de 1970 a 1995. Dividimos el periodo muestra en dos: 1970-1981 y 1982-1995; asimismo, se mostró, con base en la prueba de Chow, que existía una diferencia en la regresión del ahorro sobre el ingreso para los dos periodos.

No obstante, no pudimos determinar si dicha diferencia en las dos regresiones se debía a las diferencias en los términos del intercepto o en los coeficientes de la pendiente, o a ambas situaciones. Con mucha frecuencia, saber esto por sí mismo resulta muy útil.

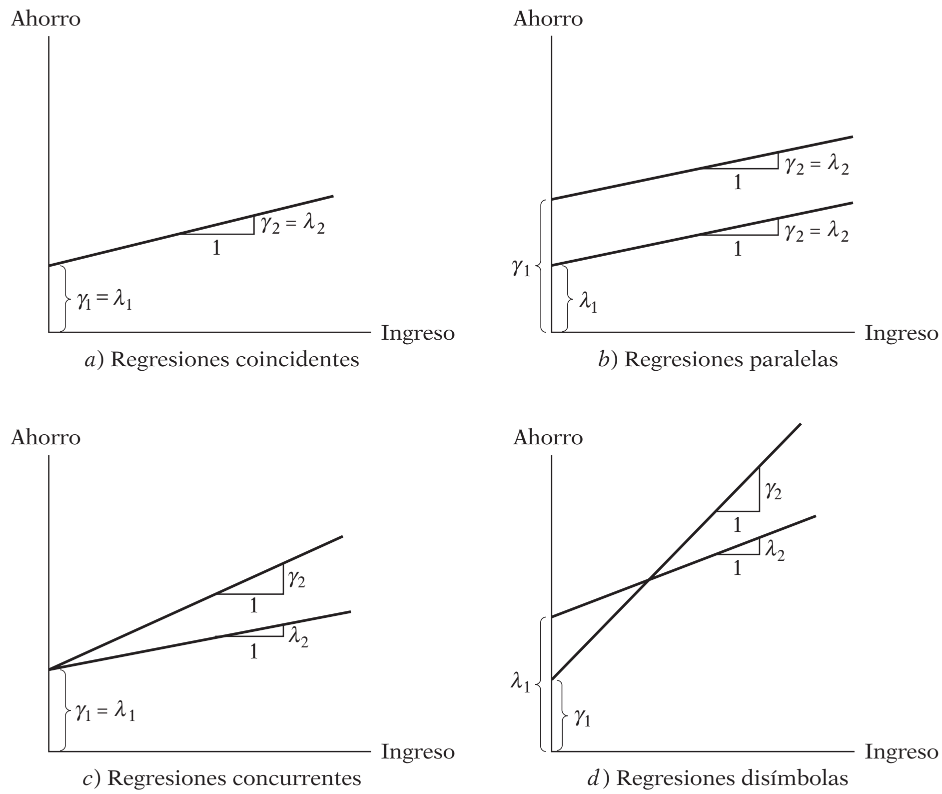
Al ver las ecuaciones (8.7.1) y (8.7.2) se observa que hay cuatro posibilidades, las cuales se ilustran en la figura 9.3:

1. El intercepto y los coeficientes de las pendientes son iguales en ambas regresiones. Esta situación, el caso de **regresiones coincidentes**, se muestra en la figura 9.3*a*.
2. Sólo los interceptos en ambas regresiones son diferentes, pero las pendientes son las mismas. Este caso, de **regresiones paralelas**, se presenta en la figura 9.3*b*.
3. Los interceptos en las dos regresiones son las mismas, pero las pendientes son distintas. Esta situación se conoce como **regresiones concurrentes** y se muestra en la figura 9.3*c*.
4. Ambos interceptos y pendientes en las dos regresiones son distintos. Este caso es el de **regresiones disímbolas**, lo cual se muestra en la figura 9.3*d*.

Como ya mencionamos, la prueba de Chow de múltiples pasos, analizada en la sección 8.7, indica sólo si dos (o más) regresiones son distintas, pero no el origen de la diferencia. Dicha

[9] El material de esta sección se basa en los artículos del autor "Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note" y "Use of Dummy Variables. . . A Generalization", ambos publicados en *American Statistician*, vol. 24, núms. 1 y 5, 1970, pp. 50-52 y 18-21.

**FIGURA 9.3**
Regresiones plausibles de ahorro-ingreso.




causa, si hay alguna, sale a la luz al agrupar todas las observaciones (26 en total) y llevar a cabo sólo una regresión múltiple, como se muestra en seguida:[10]

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \quad \textbf{(9.5.1)}$$

donde $Y =$ ahorro
$X =$ ingreso
$t =$ tiempo
$D = 1$ para las observaciones de 1982-1995
$= 0$ en otro caso (es decir, para las observaciones de 1970-1981)

La tabla 9.2 muestra la estructura de la matriz de datos.

Para ver las implicaciones de (9.5.1), y si suponemos que, como siempre, $E(u_i) = 0$, obtenemos:

*Función de ahorros medios para 1970-1981:*

$$E(Y_t \mid D_t = 0, X_t) = \alpha_1 + \beta_1 X_t \quad \textbf{(9.5.2)}$$

*Función de ahorros medios para 1982-1995:*

$$E(Y_t \mid D_t = 1, X_t) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \quad \textbf{(9.5.3)}$$

El lector notará que se trata de las mismas funciones que (8.7.1) y (8.7.2), con $\lambda_1 = \alpha_1$, $\lambda_2 = \beta_1$, $\gamma_1 = (\alpha_1 + \alpha_2)$ y $\gamma_2 = (\beta_1 + \beta_2)$. Por tanto la estimación de (9.5.1) equivale a estimar las dos funciones de ahorro individuales (8.7.1) y (8.7.2).

[10] Como en la prueba de Chow, la técnica de agrupamiento supone la homoscedasticidad; es decir, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

**TABLA 9.2**
**Datos sobre ahorro e ingreso, Estados Unidos, 1970-1995**

Fuente: *Economic Report of the President*, 1997, tabla B-28, p. 332.

| Observación | Ahorro | Ingreso | Variable dicótoma |
|---|---|---|---|
| 1970 | 61 | 727.1 | 0 |
| 1971 | 68.6 | 790.2 | 0 |
| 1972 | 63.6 | 855.3 | 0 |
| 1973 | 89.6 | 965 | 0 |
| 1974 | 97.6 | 1 054.2 | 0 |
| 1975 | 104.4 | 1 159.2 | 0 |
| 1976 | 96.4 | 1 273 | 0 |
| 1977 | 92.5 | 1 401.4 | 0 |
| 1978 | 112.6 | 1 580.1 | 0 |
| 1979 | 130.1 | 1 769.5 | 0 |
| 1980 | 161.8 | 1 973.3 | 0 |
| 1981 | 199.1 | 2 200.2 | 0 |
| 1982 | 205.5 | 2 347.3 | 1 |
| 1983 | 167 | 2 522.4 | 1 |
| 1984 | 235.7 | 2 810 | 1 |
| 1985 | 206.2 | 3 002 | 1 |
| 1986 | 196.5 | 3 187.6 | 1 |
| 1987 | 168.4 | 3 363.1 | 1 |
| 1988 | 189.1 | 3 640.8 | 1 |
| 1989 | 187.8 | 3 894.5 | 1 |
| 1990 | 208.7 | 4 166.8 | 1 |
| 1991 | 246.4 | 4 343.7 | 1 |
| 1992 | 272.6 | 4 613.7 | 1 |
| 1993 | 214.4 | 4 790.2 | 1 |
| 1994 | 189.4 | 5 021.7 | 1 |
| 1995 | 249.3 | 5 320.8 | 1 |

*Nota:* Variable dicótoma = 1 para observaciones a partir de 1982; 0 en otro caso.
Las cifras de ahorro e ingreso se expresan en miles de millones de dólares.

En (9.5.1), $\alpha_2$ es el intercepto **diferencial**, como antes; y $\beta_2$ es el **coeficiente de la pendiente diferencial** (también llamado **alterador de pendiente**), el cual indica cuánto difiere el coeficiente de la pendiente de la función ahorro del segundo periodo (la categoría que recibe el valor dicótomo de 1) respecto del primer periodo. Observe que la introducción de la variable dicótoma $D$ en la **forma interactiva**, o **multiplicativa** ($D$ multiplicada por $X$), permite diferenciar entre los coeficientes de las pendientes de los dos periodos, del mismo modo que la introducción de la variable dicótoma en **forma aditiva** permite distinguir entre los interceptos de los dos periodos.

**EJEMPLO 9.4**
*Diferencias estructurales en la regresión ahorro-ingreso para Estados Unidos: método de la variable dicótoma*

Antes de proseguir, veamos los resultados de la regresión del modelo (9.5.1) aplicada a los datos de ahorro-ingreso de Estados Unidos.

$$\begin{aligned}\hat{Y}_t &= \quad 1.0161 \quad + 152.4786D_t + \ 0.0803X_t - \ 0.0655(D_tX_t)\\ \text{ee} &= (20.1648) \quad (33.0824) \quad (0.0144) \quad (0.0159)\\ t &= \ (0.0504)^{**} \quad (4.6090)^{*} \quad (5.5413)^{*} \quad (-4.0963)^{*}\\ &\qquad\qquad\qquad\qquad\qquad R^2 = 0.8819\end{aligned} \tag{9.5.4}$$

donde * indica valores $p$ menores que 5%, y ** indica valores $p$ mayores que 5%.

(*continúa*)

**EJEMPLO 9.4**
(*continuación*)

Como muestran los resultados de esta regresión, el intercepto diferencial y el coeficiente de la pendiente son estadísticamente significativos, lo cual indica enérgicamente que las regresiones ahorro-ingreso para los dos periodos son diferentes, como en la figura 9.3*d*.

De (9.5.4) derivamos las ecuaciones (9.5.2) y (9.5.3), las cuales son:

*Regresión ahorro-ingreso para 1970-1981:*

$$\hat{Y}_t = 1.0161 + 0.0803X_t \tag{9.5.5}$$

*Regresión ahorro-ingreso para 1982-1995:*

$$\begin{aligned}\hat{Y}_t &= (1.0161 + 152.4786) + (0.0803 - 0.0655)X_t \\ &= 153.4947 + 0.0148X_t\end{aligned} \tag{9.5.6}$$

Éstos son precisamente los resultados obtenidos en (8.7.1*a*) y (8.7.2*a*), lo cual no debe sorprender. Tales regresiones ya se mostraron en la figura 8.3.

Ahora se ven de inmediato las ventajas de la técnica de la variable dicótoma [es decir, la estimación de (9.5.1)], sobre la prueba de Chow [es decir, la estimación de las tres regresiones: (8.7.1), (8.7.2) y (8.7.3)]:

1. Sólo fue necesaria una regresión, pues las regresiones individuales pueden derivarse con facilidad a partir de ella, del modo indicado por las ecuaciones (9.5.2) y (9.5.3).
2. Con la regresión (9.5.1) se prueban diversas hipótesis. Por tanto, si el coeficiente del *intercepto diferencial* $\alpha_2$ es estadísticamente insignificante, se puede aceptar la hipótesis de que las dos regresiones tienen el mismo intercepto; es decir, ambas regresiones son concurrentes (compare con la figura 9.3*c*). De modo semejante, si el coeficiente de la *pendiente diferencial* $\beta_2$ es estadísticamente insignificante pero $\alpha_2$ es significativo, tal vez no se rechace la hipótesis de que las dos regresiones tienen la misma pendiente; es decir, las dos líneas de regresión son paralelas (véase la figura 9.3*b*). La prueba de la estabilidad de toda la regresión (es decir, $\alpha_2 = \beta_2 = 0$, de manera simultánea) se lleva a cabo mediante la prueba *F* usual (recuerde la prueba *F* de los mínimos cuadrados restringidos). Si no se rechaza la hipótesis, las líneas de regresión serán coincidentes, como se aprecia en la figura 9.3*a*.
3. La prueba de Chow no establece de manera explícita *cuál* coeficiente, intercepto o pendiente es distinto, ni si ambos son diferentes en los dos periodos (como en el ejemplo anterior). Es decir, se puede tener una prueba de Chow significativa debido a que sólo la *pendiente* es diferente o a que sólo el intercepto es distinto, o porque ambos lo son. En otras palabras, no se puede saber, mediante la prueba de Chow, cuál de las cuatro posibilidades esquematizadas en la figura 9.3 es la que se tiene en una determinada instancia. Al respecto, el método de la variable dicótoma tiene una clara ventaja, pues no sólo indica si los dos periodos son distintos, sino que también destaca la(s) causa(s) de la diferencia: si se debe al intercepto, a la pendiente o a las dos. En la práctica, saber si dos regresiones difieren en uno u otro coeficiente resulta tan importante, si no más, que sólo saber que son distintas.
4. Por último, en vista de que el agrupamiento (es decir, incluir todas las observaciones en una sola regresión) aumenta los grados de libertad, tal vez mejore la precisión relativa de los parámetros estimados. Por supuesto, tenga en cuenta que cada inclusión de una variable dicótoma consumirá un grado de libertad.

## 9.6 Efectos de interacción al utilizar variables dicótomas

Las variables dicótomas son una herramienta flexible para varios problemas interesantes. Observemos lo anterior con el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \tag{9.6.1}$$

donde $Y =$ salario por hora en dólares
$X =$ educación (años de escolaridad)
$D_2 =$ 1 si es mujer; 0 en otro caso
$D_3 =$ 1 si no es blanco y no hispano; 0 en otro caso

En este modelo, el sexo y la raza son regresoras cualitativas y la escolaridad es cuantitativa.[11] Está implícito en este modelo el supuesto de que el efecto diferencial de la variable dicótoma sexo, $D_2$, es constante en las dos categorías de raza, y el efecto diferencial de la variable dicótoma raza, $D_3$, también es constante en ambos sexos. Es decir, si el salario medio es mayor para los hombres que para las mujeres, esto ocurre independientemente de que sean no blancos/no hispanos o no. De igual forma, si por ejemplo los no blancos/no hispanos tienen salarios medios menores, esto ocurre independientemente de que sean hombres o mujeres.

En muchas aplicaciones dicho supuesto puede ser insostenible. Una mujer no blanca ni hispana tal vez gane menor salario que un hombre de esa misma categoría. En otras palabras, quizá haya **interacción** entre las dos variables cualitativas $D_2$ y $D_3$. Por tanto, su efecto sobre la media $Y$ quizá no sea simplemente **aditivo**, como en (9.6.1), sino también **multiplicativo**, como en el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \tag{9.6.2}$$

donde las variables están definidas como en el modelo (9.6.1).

De (9.6.2) obtenemos

$$E(Y_i \mid D_{2i} = 1, D_{3i} = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i \tag{9.6.3}$$

que es la función salario medio por hora para las trabajadoras no blancas ni hispanas. Observe que

$\alpha_2 =$ efecto diferencial de ser mujer
$\alpha_3 =$ efecto diferencial de ser no blanco ni hispano
$\alpha_4 =$ efecto diferencial de ser mujer no blanca ni hispana

lo cual muestra que el salario medio por hora de las mujeres no blancas ni hispanas es diferente (en una cantidad igual a $\alpha_4$) del salario medio por hora de las mujeres blancas o hispanas. Si por ejemplo los tres coeficientes de las variables dicótomas son negativos, se implica que las trabajadoras no blancas ni hispanas ganan un salario medio por hora mucho más bajo que las trabajadoras blancas o hispanas, en comparación con la categoría base, la cual en el ejemplo presente es la de hombres blancos o hispanos.

Ahora el lector puede observar la forma en que la **variable dicótoma de interacción** (es decir, el producto de dos variables cualitativas o dicótomas) modifica el efecto de los dos atributos considerados de manera individual (es decir, en forma aditiva).

**EJEMPLO 9.5**
*Ingreso promedio por hora en comparación con la escolaridad, sexo y raza*

Veamos primero los resultados de la regresión basados en el modelo (9.6.1). Con los datos con que se estimó la regresión (9.3.1) obtuvimos lo siguiente:

$$\begin{aligned} \hat{Y}_i &= -0.2610 \quad - \quad 2.3606 D_{2i} - \quad 1.7327 D_{3i} + \; 0.8028 X_i \\ t &= (-0.2357)^{**} \quad (-5.4873)^{*} \quad (-2.1803)^{*} \quad (9.9094)^{*} \\ & \qquad R^2 = 0.2032 \qquad n = 528 \end{aligned} \tag{9.6.4}$$

donde * indica valores $p$ menores que 5% y ** indica valores $p$ mayores que 5%.

(*continúa*)

[11] Si definiéramos la variable escolaridad como menos que educación media superior, educación media superior y más que educación media superior, podríamos utilizar entonces dos variables dicótomas para representar las tres clases.

**EJEMPLO 9.5** (*continuación*)

El lector puede verificar que los coeficientes de intercepto diferenciales son estadísticamente significativos, que tienen los signos que se esperaban (¿por qué?) y que la escolaridad tiene un gran efecto positivo sobre el salario por hora, lo cual no causa sorpresa alguna.

Como lo muestra (9.6.4), *ceteris paribus*, el ingreso promedio por hora de las mujeres es inferior por cerca de $2.36; además, el ingreso promedio por hora de los trabajadores no blancos ni hispanos también es menor por aproximadamente $1.73.

Ahora consideremos los resultados del modelo (9.6.2), que incluyen la variable dicótoma de interacción.

$$\begin{aligned} \hat{Y}_i &= -0.26100 \quad - \quad 2.3606D_{2i} - \quad 1.7327D_{3i} + 2.1289D_{2i}D_{3i} + 0.8028X_i \\ t &= (-0.2357)^{**} \quad (-5.4873)^{*} \quad (-2.1803)^{*} \quad (1.7420)^{**} \quad (9.9095)^{**} \\ & \qquad R^2 = 0.2032 \qquad n = 528 \end{aligned} \tag{9.6.5}$$

donde * indica valores $p$ menores que 5% y ** indica valores $p$ mayores que 5%.

Como se nota, las dos variables dicótomas aditivas son aún estadísticamente significativas, pero la variable dicótoma interactiva no está al nivel convencional de 5%; el valor $p$ real de la variable de interacción es de un nivel de casi 8%. Si consideramos que esto es una probabilidad lo bastante baja, interpretamos los resultados de (9.6.5) de la siguiente manera: si se mantiene constante el nivel de educación y se suman los tres coeficientes de las variables dicótomas, obtendremos $-1.964$ ($= -2.3605 - 1.7327 + 2.1289$), lo cual significa que los salarios medios por hora de las trabajadoras no blancas ni hispanas es menor por casi $1.96, valor que está entre $-2.3605$ (diferencia debido sólo al sexo) y $-1.7327$ (diferencia debida sólo a la raza).

El ejemplo anterior revela con claridad la interacción de las variables dicótomas cuando se incluyen en el modelo dos o más regresoras cualitativas. Es importante observar que en el modelo (9.6.5) se supone que la tasa de crecimiento de los ingresos por hora respecto de la escolaridad (de cerca de 80 centavos por año adicional de escolaridad) sigue constante en cuanto al sexo y la raza. Pero esto tal vez no sea así. Si desea probar lo anterior deberá introducir coeficientes de pendiente diferenciales (véase el ejercicio 9.25).

## 9.7 Uso de las variables dicótomas en el análisis estacional

Muchas series de tiempo económicas que se basan en datos mensuales o trimestrales presentan pautas estacionales (movimiento oscilatorio regular); por ejemplo, las ventas de las tiendas de departamentos en la época de Navidad y otras festividades importantes, la demanda de dinero (saldos de efectivo) por parte de las familias en épocas de vacaciones, la demanda de helado y bebidas gaseosas durante el verano y los precios de los cultivos justo después de la época de cosecha, la demanda de viajes en avión, etc. A menudo es útil eliminar el factor o *componente* estacional de las series de tiempo con el fin de concentrarse en los demás componentes, como la tendencia.[12] El proceso de eliminar el componente estacional de una serie de tiempo se conoce como **desestacionalización** o **ajuste estacional**, y la serie de tiempo así obtenida se denomina serie de tiempo **desestacionalizada** o **ajustada por estacionalidad**. Las series de tiempo económicas importantes, como el índice de precios al consumidor (IPC), el índice de precios del productor (IPP) y el índice de producción industrial, suelen publicarse ajustadas por estacionalidad.

[12] Una serie de tiempo puede tener cuatro componentes: (1) **estacional**, (2) **cíclico**, (3) **tendencia** y (4) **estrictamente aleatorio**.

**TABLA 9.3**
**Datos trimestrales sobre ventas de aparatos electrodomésticos (en miles) y gasto en bienes duraderos (del primer trimestre de 1978 al cuarto de 1985)**

Fuente: *Business Statistics and Survey of Current Business*, Department of Commerce (varios números).

| LAVLZ | TRIT | REFR | LAVD | BIDU | LAVLZ | TRIT | REFR | LAVD | BIDU |
|---|---|---|---|---|---|---|---|---|---|
| 841 | 798 | 1 317 | 1 271 | 252.6 | 480 | 706 | 943 | 1 036 | 247.7 |
| 957 | 837 | 1 615 | 1 295 | 272.4 | 530 | 582 | 1 175 | 1 019 | 249.1 |
| 999 | 821 | 1 662 | 1 313 | 270.9 | 557 | 659 | 1 269 | 1 047 | 251.8 |
| 960 | 858 | 1 295 | 1 150 | 273.9 | 602 | 837 | 973 | 918 | 262 |
| 894 | 837 | 1 271 | 1 289 | 268.9 | 658 | 867 | 1 102 | 1 137 | 263.3 |
| 851 | 838 | 1 555 | 1 245 | 262.9 | 749 | 860 | 1 344 | 1 167 | 280 |
| 863 | 832 | 1 639 | 1 270 | 270.9 | 827 | 918 | 1 641 | 1 230 | 288.5 |
| 878 | 818 | 1 238 | 1 103 | 263.4 | 858 | 1 017 | 1 225 | 1 081 | 300.5 |
| 792 | 868 | 1 277 | 1 273 | 260.6 | 808 | 1 063 | 1 429 | 1 326 | 312.6 |
| 589 | 623 | 1 258 | 1 031 | 231.9 | 840 | 955 | 1 699 | 1 228 | 322.5 |
| 657 | 662 | 1 417 | 1 143 | 242.7 | 893 | 973 | 1 749 | 1 297 | 324.3 |
| 699 | 822 | 1 185 | 1 101 | 248.6 | 950 | 1 096 | 1 117 | 1 198 | 333.1 |
| 675 | 871 | 1 196 | 1 181 | 258.7 | 838 | 1 086 | 1 242 | 1 292 | 344.8 |
| 652 | 791 | 1 410 | 1 116 | 248.4 | 884 | 990 | 1 684 | 1 342 | 350.3 |
| 628 | 759 | 1 417 | 1 190 | 255.5 | 905 | 1 028 | 1 764 | 1 323 | 369.1 |
| 529 | 734 | 919 | 1 125 | 240.4 | 909 | 1 003 | 1 328 | 1 274 | 356.4 |

*Nota:* LAVLZ = lavalozas; TRIT = trituradores de basura; REFR = refrigeradores; LAVD = lavadoras; BIDU = gasto en bienes duraderos, miles de millones de dólares de 1982.

Hay diversos métodos para desestacionalizar una serie de tiempo, pero consideraremos sólo uno: el *método de las variables dicótomas*.[13] Para ilustrar la forma de desestacionalizar con las variables dicótomas, considere los datos de la tabla 9.3. Se trata de datos trimestrales de 1978 a 1995 respecto de las ventas de cuatro aparatos principales: lavalozas, trituradores de basura, refrigeradores y lavadoras, en miles de unidades. La tabla también suministra datos sobre el gasto en bienes duraderos en 1982, en miles de millones de dólares.

A fin de ilustrar la técnica de la variable dicótoma consideraremos sólo las ventas de los refrigeradores en el periodo de muestra. Pero primero observe los datos de la figura 9.4. Esa figura indica que tal vez exista un modelo estacional en los datos asociados con los diversos trimestres. Para verificarlo, vea el siguiente modelo:

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_{3t} D_{3t} + \alpha_4 D_{4t} + u_t \qquad \textbf{(9.7.1)}$$

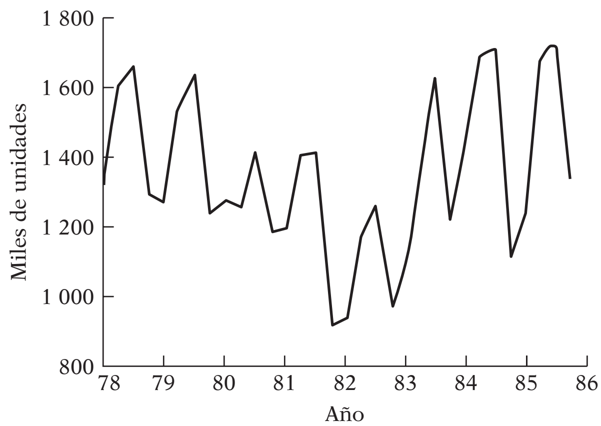
donde $Y_t$ = ventas de refrigeradores (en miles) y las $D$ son las variables dicótomas, las cuales toman un valor de 1 en el trimestre relevante, y 0 en otro caso. *Observe que para evitar la trampa de la variable dicótoma asignamos una variable dicótoma a cada trimestre del año, pero omitimos el término del intercepto.* Si hubiera algún efecto estacional en un determinado trimestre se señalaría mediante un valor $t$ estadísticamente significativo del coeficiente de la variable dicótoma para dicho trimestre.[14]

Observe que en (9.7.1) se hace la regresión de $Y$ efectivamente sobre un intercepto, salvo que se permite un intercepto distinto para cada temporada (es decir, trimestre). Como resultado, el coeficiente de la variable dicótoma de cada trimestre proporcionará la media de las ventas de refrigeradores de cada trimestre o temporada (¿por qué?).

[13] Para los diversos métodos de ajuste estacional, véase por ejemplo Francis X. Diebold, *Elements of Forecasting*, 2a. ed., South-Western Publishers, 2001, capítulo 5.

[14] Considere un aspecto técnico. Este método de asignar una variable dicótoma a cada trimestre supone que el factor estacional, si está presente, es determinista y no estocástico. Volveremos a este tema cuando analicemos la econometría de las series de tiempo, en la parte V de este libro.

**FIGURA 9.4**
Ventas de refrigeradores, 1978-1985 (trimestrales).




**EJEMPLO 9.6**
*Estacionalidad en las ventas de refrigeradores*

De los datos sobre las ventas de refrigeradores de la tabla 9.4 se obtienen los siguientes resultados de la regresión:

$$\begin{aligned}\hat{Y}_t &= 1\,222.125D_{1t} + 1\,467.500D_{2t} + 1\,569.750D_{3t} + 1\,160.000D_{4t}\\ t &= \quad (20.3720) \qquad (24.4622) \qquad (26.1666) \qquad (19.3364)\\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad R^2 = 0.5317\end{aligned} \tag{9.7.2}$$

*Nota:* No se proporcionaron los errores estándar de los coeficientes estimados: cada uno de ellos es igual a 59.9904, pues todas las variables dicótomas sólo toman el valor de 1 o de 0.

Los coeficientes estimados $\alpha$ en (9.7.2) representan el promedio, o *media*, de las ventas de refrigeradores (en miles de unidades) en cada temporada (es decir, trimestre). Por tanto, el promedio de ventas de refrigeradores en el primer trimestre, en miles de unidades, es de casi 1 222, en el segundo trimestre fue de casi 1 468, las del tercer trimestre fueron de 1 570 aproximadamente, y las del último trimestre fueron de casi 1 160.

**TABLA 9.4**
**Ventas de refrigeradores (miles) en Estados Unidos, de 1978 a 1985 (trimestrales)**

Fuente: *Business Statistics and Survey of Current Business*, Department of Commerce (varios números).

| REFRI | BIDU | $D_2$ | $D_3$ | $D_4$ | REFRI | BIDU | $D_2$ | $D_3$ | $D_4$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 317 | 252.6 | 0 | 0 | 0 | 943 | 247.7 | 0 | 0 | 0 |
| 1 615 | 272.4 | 1 | 0 | 0 | 1 175 | 249.1 | 1 | 0 | 0 |
| 1 662 | 270.9 | 0 | 1 | 0 | 1 269 | 251.8 | 0 | 1 | 0 |
| 1 295 | 273.9 | 0 | 0 | 1 | 973 | 262.0 | 0 | 0 | 1 |
| 1 271 | 268.9 | 0 | 0 | 0 | 1 102 | 263.3 | 0 | 0 | 0 |
| 1 555 | 262.9 | 1 | 0 | 0 | 1 344 | 280.0 | 1 | 0 | 0 |
| 1 639 | 270.9 | 0 | 1 | 0 | 1 641 | 288.5 | 0 | 1 | 0 |
| 1 238 | 263.4 | 0 | 0 | 1 | 1 225 | 300.5 | 0 | 0 | 1 |
| 1 277 | 260.6 | 0 | 0 | 0 | 1 429 | 312.6 | 0 | 0 | 0 |
| 1 258 | 231.9 | 1 | 0 | 0 | 1 699 | 322.5 | 1 | 0 | 0 |
| 1 417 | 242.7 | 0 | 1 | 0 | 1 749 | 324.3 | 0 | 1 | 0 |
| 1 185 | 248.6 | 0 | 0 | 1 | 1 117 | 333.1 | 0 | 0 | 1 |
| 1 196 | 258.7 | 0 | 0 | 0 | 1 242 | 344.8 | 0 | 0 | 0 |
| 1 410 | 248.4 | 1 | 0 | 0 | 1 684 | 350.3 | 1 | 0 | 0 |
| 1 417 | 255.5 | 0 | 1 | 0 | 1 764 | 369.1 | 0 | 1 | 0 |
| 919 | 240.4 | 0 | 0 | 1 | 1 328 | 356.4 | 0 | 0 | 1 |

*Nota:* REFRI = ventas de refrigeradores, miles de unidades.
BIDU = gasto en bienes duraderos, miles de millones de dólares de 1982.
$D_2$ = 1 en el segundo trimestre; 0 en otro caso.
$D_3$ = 1 en el tercer trimestre; 0 en otro caso.
$D_4$ = 1 en el cuarto trimestre; 0 en otro caso.

**EJEMPLO 9.6** (*continuación*)

A propósito, en vez de asignar una variable dicótoma a cada trimestre y suprimir el término del intercepto a fin de evitar la trampa de variable dicótoma, se puede asignar sólo tres variables dicótomas e incluir el término del intercepto. Suponga que consideramos el primer trimestre como referencia y asignamos variables dicótomas al segundo, tercero y cuarto. Lo anterior da los siguientes resultados de regresión (véase la tabla 9.4 para la organización de los datos):

$$\begin{aligned} \hat{Y}_t &= 1\,222.1250 \quad + 245.3750D_{2t} + 347.6250D_{3t} - \quad 62.1250D_{4t} \\ t &= \quad (20.3720)^* \quad (2.8922)^* \quad (4.0974)^* \quad (-0.7322)^{**} \\ & \qquad\qquad R^2 = 0.5318 \end{aligned} \tag{9.7.3}$$

donde * indica valores *p* menores que 5%, y ** indica valores *p* mayores que 5%.

Como consideramos el primer trimestre como punto de referencia, los coeficientes relacionados con las distintas variables dicótomas ahora son *interceptos diferenciales* que muestran en qué medida el *valor promedio* de $Y$ en el trimestre que recibe un valor de 1 para la variable dicótoma difiere del trimestre que es punto de referencia. En otras palabras, los coeficientes de las variables estacionales indican el incremento o decremento estacional del valor promedio de $Y$ en relación con la temporada base. Si se suman los distintos valores del intercepto diferencial al valor promedio de referencia de 1 222.125, se tendrá el valor promedio para los distintos trimestres. Al llevar a cabo lo anterior se reproducirá exactamente la ecuación (9.7.2), salvo errores de redondeo.

Ahora apreciará el valor de considerar un trimestre como punto de referencia, pues (9.7.3) muestra que el valor promedio de $Y$ para el cuarto trimestre no es estadísticamente distinto del valor promedio para el primer trimestre, porque el coeficiente de la variable dicótoma para el cuarto trimestre no es estadísticamente significativo. Por supuesto, la respuesta cambia según el trimestre con que se compare; no obstante, la conclusión general sigue siendo la misma.

¿Cómo obtener la serie de tiempo desestacionalizada de las ventas de los refrigeradores? Es fácil. Se estiman los valores $Y$ a partir del modelo (9.7.2) [o (9.7.3)] para cada observación y se restan de los valores reales de $Y$; es decir, se obtiene $(Y_t - \hat{Y}_t)$, que son sólo los residuos de la regresión (9.7.2), los cuales se presentan en la tabla 9.5.[15] A estos residuos es necesario sumarles la media de las series $Y$ para obtener los valores pronosticados.

¿Qué representan estos residuos? Significan los componentes que quedan de la serie de tiempo de los refrigeradores, a saber, la tendencia, el ciclo y el componente aleatorio (pero tenga en cuenta la advertencia de la nota 15).

Como los modelos (9.7.2) y (9.7.3) no contienen covariantes, ¿cambiaría la situación si se añade una regresora cuantitativa al modelo? Por la influencia del gasto en bienes duraderos sobre la demanda de refrigeradores, el modelo (9.7.3) se extenderá para incluir esta variable. Los datos para el gasto en bienes duraderos en miles de millones de dólares de 1982 ya se proporcionaron en la tabla 9.3. Ésta es la variable $X$ (cuantitativa) del modelo. Los resultados de la regresión son los siguientes:

$$\begin{aligned} \hat{Y}_t &= 456.2440 \quad + 242.4976D_{2t} + 325.2643D_{3t} - \quad 86.0804D_{4t} + \; 2.7734X_t \\ t &= \quad (2.5593)^* \quad (3.6951)^* \quad (4.9421)^* \quad (-1.3073)^{**} \quad (4.4496)^* \\ & \qquad\qquad R^2 = 0.7298 \end{aligned} \tag{9.7.4}$$

donde * indica valores *p* menores que 5% y ** indica valores *p* mayores que 5%.

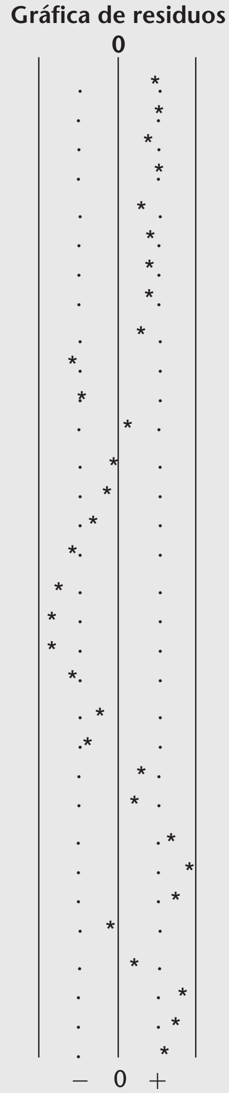
(*continúa*)

[15] Desde luego, esto supone que la técnica de las variables dicótomas es adecuada para desestacionalizar una serie de tiempo, y que una serie de tiempo (ST) puede representarse como ST $= s + c + t + u$, donde $s$ indica la estacionalidad, $t$ la tendencia, $c$ el ciclo y $u$ el componente aleatorio. No obstante, si la serie de tiempo es de la forma ST $= (s)(c)(t)(u)$, donde las cuatro componentes ingresan de manera multiplicativa, el método anterior para desestacionalizar resulta inapropiado, pues supone que los cuatro componentes de una serie de tiempo son aditivos. Sin embargo, diremos más al respecto en los capítulos sobre econometría de las series de tiempo.

**EJEMPLO 9.6** (*continuación*)

TABLA 9.5 **Regresión de las ventas de refrigeradores: valores reales, ajustados y residuales (ec. 9.7.3)**

| | Real | Ajustado | Residual |
|---|---|---|---|
| 1978-I | 1 317 | 1 222.12 | 94.875 |
| 1978-II | 1 615 | 1 467.50 | 147.500 |
| 1978-III | 1 662 | 1 569.75 | 92.250 |
| 1978-IV | 1 295 | 1 160.00 | 135.000 |
| 1979-I | 1 271 | 1 222.12 | 48.875 |
| 1979-II | 1 555 | 1 467.50 | 87.500 |
| 1979-III | 1 639 | 1 569.75 | 69.250 |
| 1979-IV | 1 238 | 1 160.00 | 78.000 |
| 1980-I | 1 277 | 1 222.12 | 54.875 |
| 1980-II | 1 258 | 1 467.50 | −209.500 |
| 1980-III | 1 417 | 1 569.75 | −152.750 |
| 1980-IV | 1 185 | 1 160.00 | 25.000 |
| 1981-I | 1 196 | 1 222.12 | −26.125 |
| 1981-II | 1 410 | 1 467.50 | −57.500 |
| 1981-III | 1 417 | 1 569.75 | −152.750 |
| 1981-IV | 919 | 1 160.00 | −241.000 |
| 1982-I | 943 | 1 222.12 | −279.125 |
| 1982-II | 1 175 | 1 467.50 | −292.500 |
| 1982-III | 1 269 | 1 569.75 | −300.750 |
| 1982-IV | 973 | 1 160.00 | −187.000 |
| 1983-I | 1 102 | 1 222.12 | −120.125 |
| 1983-II | 1 344 | 1 467.50 | −123.500 |
| 1983-III | 1 641 | 1 569.75 | 71.250 |
| 1983-IV | 1 225 | 1 160.00 | 65.000 |
| 1984-I | 1 429 | 1 222.12 | 206.875 |
| 1984-II | 1 699 | 1 467.50 | 231.500 |
| 1984-III | 1 749 | 1 569.75 | 179.250 |
| 1984-IV | 1 117 | 1 160.00 | −43.000 |
| 1985-I | 1 242 | 1 222.12 | 19.875 |
| 1985-II | 1 684 | 1 467.50 | 216.500 |
| 1985-III | 1 764 | 1 569.75 | 194.250 |
| 1985-IV | 1 328 | 1 160.00 | 168.000 |




De nuevo, tenga en cuenta que consideramos como base al primer trimestre. Al igual que en (9.7.3), vea que los coeficientes de intercepto diferenciales para el segundo y tercer trimestres son estadísticamente diferentes de los del primer trimestre; pero los interceptos del cuarto y primer trimestres son estadísticamente iguales. El coeficiente de *X* (gasto en bienes duraderos), cercano a 2.77, indica que, si se permiten los efectos estacionales y el gasto en bienes duraderos se incrementa un dólar las ventas de refrigeradores aumentan en promedio cerca de 2.77 unidades; es decir, aproximadamente 3 unidades. Considere que los refrigeradores están dados en miles de unidades y *X* está en miles de millones de dólares (de 1982).

Una pregunta interesante es: al igual que las ventas de refrigeradores muestran patrones estacionales, ¿el gasto en bienes duraderos también presenta patrones estacionales? ¿Cómo tomar en cuenta entonces la naturaleza estacional de *X*? Lo destacable respecto de (9.7.4) es que las variables dicótomas de ese modelo no sólo eliminan la estacionalidad en *Y*, sino que también la estacionalidad, si existe, en *X*. (Esto se deduce del **teorema de Frisch-Waugh**, muy conocido

**EJEMPLO 9.6** (*continuación*)

en estadística.)[16] Por decirlo así, se matan (desestacionalizan) dos pájaros (dos series) con un tiro (la técnica de variable dicótoma).

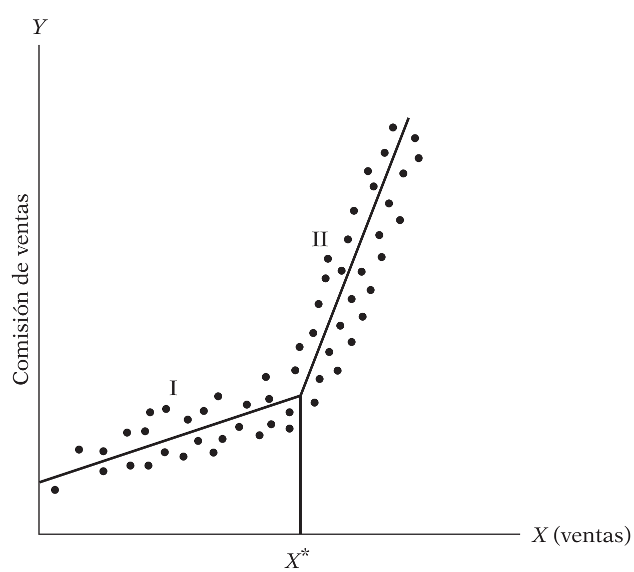
Si desea una prueba informal del enunciado anterior, sólo siga estos pasos: 1) Haga la regresión de $Y$ sobre las variables dicótomas, como en (9.7.2) o (9.7.3), y guarde los residuos, por ejemplo $S_1$; representan a la $Y$ desestacionalizada. 2) Efectúe una regresión similar para $X$ y obtenga los residuos de esta regresión, por ejemplo $S_2$; tales residuos representan a la $X$ desestacionalizada. 3) Lleve a cabo la regresión de $S_1$ sobre $S_2$. Descubrirá que el coeficiente de la pendiente de esta regresión es precisamente el coeficiente de $X$ de la regresión (9.7.4).

## 9.8 Regresión lineal por segmentos

Para ilustrar una vez más el uso de las variables dicótomas, considere la figura 9.5, que muestra la forma como una compañía hipotética remunera a sus representantes de ventas. Las comisiones se pagan con base en las ventas de forma que, hasta un cierto nivel, *meta* o *umbral*, nivel $X^*$, existe una estructura de comisiones (estocástica), mientras que por encima de ese nivel existe otra. (*Nota:* Además de las ventas, hay otros factores que afectan la comisión de las ventas. Suponga que estos otros factores están representados por el término de perturbación estocástico.) Más específicamente, se supone que la comisión de ventas aumenta linealmente con las ventas hasta el nivel del umbral $X^*$, después del cual ésta también aumenta linealmente con las ventas pero a una tasa mayor. Por tanto, se tiene una **regresión lineal por segmentos** que consta de dos partes o segmentos lineales, a los cuales se les da el nombre de I y II en la figura 9.5, y la función de las comisiones cambia su pendiente en el valor del umbral. Dados los datos sobre comisiones, ventas y el valor del nivel del umbral $X^*$, con la técnica de las variables dicótomas se estiman las diferentes pendientes de los dos segmentos de la regresión lineal por secciones que aparece en la figura 9.5. Procedemos de la siguiente manera:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i \qquad \textbf{(9.8.1)}$$

**FIGURA 9.5**
Relación hipotética entre las comisiones de ventas y el volumen de ventas. (*Nota:* El intercepto en el eje $Y$ denota una comisión mínima garantizada.)




[16] Para una demostración, véase Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, Lyme, Gran Bretaña, 1995, pp. 150-152.

donde $Y_i =$ comisión de ventas
$X_i =$ volumen de ventas generado por el vendedor
$X^* =$ valor del umbral de las ventas, conocido también como **nudo** (conocido por anticipado)[17]
$D = 1$ si $X_i > X^*$
$= 0$ si $X_i < X^*$

Si $E(u_i) = 0$ vemos inmediatamente que

$$E(Y_i \mid D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i \tag{9.8.2}$$

el cual muestra la comisión de ventas promedio hasta el nivel objetivo $X^*$ y

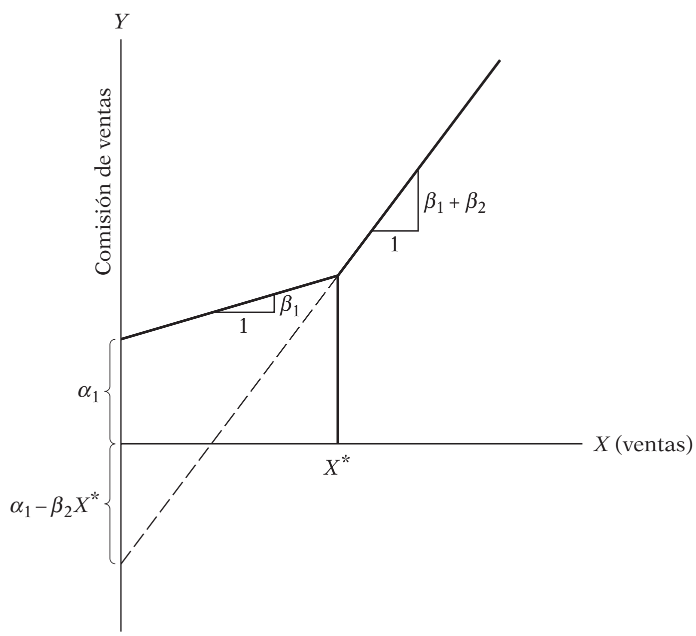
$$E(Y_i \mid D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_2) X_i \tag{9.8.3}$$

que muestra la comisión de ventas promedio más allá del nivel objetivo $X^*$.

Así, $\beta_1$ corresponde a la pendiente de la línea de regresión en el segmento I y $\beta_1 + \beta_2$ corresponde a la pendiente de la línea de regresión en el segmento II de la regresión lineal por segmentos de la figura 9.5. Es fácil probar la hipótesis de que no existe, en la regresión, una discontinuidad en el valor del umbral $X^*$ al observar la significancia estadística del coeficiente de pendiente diferencial estimado $\hat{\beta}_2$ (véase la figura 9.6).

A propósito, la regresión lineal por segmentos que acabamos de exponer ejemplifica una clase más general de funciones conocidas como **funciones "spline"**.[18]

**FIGURA 9.6**
Parámetros de la regresión lineal por segmentos.




[17] Sin embargo, tal vez no siempre sea fácil identificar el valor del umbral. Un enfoque *ad hoc* consiste en graficar la variable dependiente frente a la(s) variable(s) explicativa(s) y observar si parece haber un cambio pronunciado en la relación después de un valor dado de $X$ (por ejemplo, $X^*$). Un enfoque analítico para hallar el punto de ruptura se encuentra en los llamados **modelos de regresión cambiantes**. Pero se trata un tema avanzado, y acerca del cual hay un análisis clásico en Thomas Fomby, R. Carter Hill y Stanley Johnson, *Advanced Econometric Methods*, Springer-Verlag, Nueva York, 1984, capítulo 14.

[18] Hay una exposición sencilla de las funciones "spline" (por ejemplo, los polinomios por segmentos de orden *k*) en Douglas C. Montgomery y Elizabeth A. Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, 3a. ed., Nueva York, 2001, pp. 228-230.

**EJEMPLO 9.7**
*Costo total en relación con la producción*

Como ejemplo de la aplicación de la regresión lineal por segmentos, considere los datos hipotéticos de costo total-producción total presentados en la tabla 9.6. Se dice que el costo total puede cambiar su pendiente al alcanzar un nivel de producción de 5 500 unidades.

Si $Y$ en (9.8.4) representa el costo total y $X$ la producción total, obtenemos los siguientes resultados:

$$\begin{aligned} \hat{Y}_i &= -145.72 \quad + \; 0.2791X_i + \; 0.0945(X_i - X_i^*)D_i \\ t &= \quad (-0.8245) \quad (6.0669) \quad (1.1447) \\ & \qquad R^2 = 0.9737 \qquad X^* = 5\,500 \end{aligned} \tag{9.8.4}$$

Como muestran estos resultados, el costo marginal de producción es de cerca de 28 centavos de dólar por unidad, y aunque éste es de cerca de 37 centavos (28 + 9) para la producción por encima de 5 500 unidades, la diferencia entre ambos no es estadísticamente significativa, pues la variable dicótoma no es significativa, por ejemplo, en el nivel de 5%. Para todos los fines prácticos, entonces, podemos efectuar la regresión del costo total sobre la producción total al eliminar la variable dicótoma.

**TABLA 9.6**
**Datos hipotéticos sobre producción y costo total**

| Costo total, dólares | Unidades de producción |
|---|---|
| 256 | 1 000 |
| 414 | 2 000 |
| 634 | 3 000 |
| 778 | 4 000 |
| 1 003 | 5 000 |
| 1 839 | 6 000 |
| 2 081 | 7 000 |
| 2 423 | 8 000 |
| 2 734 | 9 000 |
| 2 914 | 10 000 |

## 9.9 Modelos de regresión con datos en panel

Recuerde que en el capítulo 1 analizamos una serie de datos disponibles para el análisis empírico, como los *transversales*, las *series de tiempo*, los *agrupados* (una combinación de series de tiempo y datos transversales) y los *datos en panel*. La técnica de la variable dicótoma se extiende sin problemas a los datos agrupados y en panel. Como los datos en panel son cada vez más populares en el trabajo aplicado, analizaremos este tema con más detalle en el capítulo 16.

## 9.10 Algunos aspectos técnicos de la técnica con variables dicótomas

### Interpretación de variables dicótomas en regresiones semilogarítmicas

En el capítulo 6 vimos los modelos log-lin, donde la regresada es logarítmica y las regresoras son lineales. En tales modelos, los coeficientes de las pendientes de las regresoras indican la *semi*elasticidad, o el cambio porcentual en la regresada debido a una unidad de cambio en la regre-

sora: *esto sólo se cumple si la regresora es cuantitativa.* ¿Qué sucede si una regresora es una variable dicótoma? Para ser específicos, considere el siguiente modelo:

$$\ln Y_i = \beta_1 + \beta_2 D_i + u_i \tag{9.10.1}$$

donde $Y$ = tasa de salario por hora ($) y $D = 1$ para mujer y 0 para hombre.

¿Cómo interpretamos un modelo así? Si suponemos que $E(u_i) = 0$ obtenemos:

*Función salario para trabajadores:*

$$E(\ln Y_i \mid D_i = 0) = \beta_1 \tag{9.10.2}$$

*Función salario para trabajadoras:*

$$E(\ln Y_i \mid D_i = 1) = \beta_1 + \beta_2 \tag{9.10.3}$$

Por tanto, el intercepto $\beta_1$ proporciona el *logaritmo de los ingresos medios por hora,* y el coeficiente de la "pendiente", la diferencia entre el logaritmo de los ingresos medios por hora entre hombres y mujeres. Lo anterior representa una forma más bien extraña de enunciar las cosas. Pero si tomamos el antilogaritmo de $\beta_1$, lo que ahora tenemos *no* son los salarios medios por hora de los trabajadores, sino la **mediana** de los salarios. Como se sabe, *media*, *mediana* y *moda* son las tres medidas de tendencia central de una variable aleatoria. Y si tomamos el antilogaritmo de $(\beta_1 + \beta_2)$, obtenemos la mediana de los salarios por hora de las trabajadoras.

**EJEMPLO 9.8**
*Logaritmo de salarios por hora en relación con el sexo*

Para ilustrar (9.10.1) empleamos los datos implícitos en el ejemplo 9.2. Los resultados de la regresión basada en 528 observaciones son los siguientes:

$$\begin{aligned} \widehat{\ln Y_i} &= \quad 2.1763 \quad - \quad 0.2437D_i \\ t &= (72.2943)^* \quad (-5.5048)^* \\ & \qquad R^2 = 0.0544 \end{aligned} \tag{9.10.4}$$

donde * indica valores $p$ que son prácticamente cero.

Con el antilogaritmo de 2.1763 tenemos 8.8136 ($), que es la mediana de los ingresos por hora de los trabajadores, y si tomamos el antilogaritmo de [(2.1763 − 0.2437) = 1.92857], obtenemos 6.8796 ($), que es la mediana de los ingresos por hora de las trabajadoras. Por tanto, la mediana de los ingresos por hora de las trabajadoras es menor por casi 21.94% en comparación con sus contrapartes masculinos [(8.8136 − 6.8796)/8.8136].

Vale la pena notar que es posible obtener la semielasticidad para una regresora dicótoma de manera directa, mediante el proceso propuesto por Halvorsen y Palmquist.[19] *Tome el antilogaritmo (base e) del coeficiente estimado de la variable dicótoma, reste 1 y multiplique la diferencia por 100*. (Abordamos la lógica de este procedimiento en el apéndice 9.A.1.) En consecuencia, si tomamos el antilogaritmo de −0.2437 obtendremos 0.78366. Al restar 1 de lo anterior tenemos −0.2163, y después de multiplicar esta cifra por 100, −21.63%, lo cual indica que la mediana del salario de una trabajadora ($D = 1$) es menor que la de su contraparte masculina por aproximadamente 21.63%, lo mismo que obtuvimos antes, salvo errores de redondeo.

## Variables dicótomas y heteroscedasticidad

Examinemos de nuevo la regresión de ahorro-ingreso para Estados Unidos de 1970 a 1981 y 1982 a 1995, así como, en conjunto, de 1970 a 1995. Al probar la estabilidad estructural mediante la técnica de la variable dicótoma, supusimos que la varianza del error var $(u_{1i})$ = var $(u_{2i}) = \sigma^2$,

[19] Robert Halvorsen y Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations", *American Economic Review*, vol. 70, núm. 3, pp. 474-475.

es decir, que las varianzas del error en los dos periodos eran las mismas. También supusimos lo anterior para la prueba de Chow. Si tal supuesto no es válido —es decir, si las varianzas del error en los dos subperiodos son distintas— es muy probable que se hagan deducciones incorrectas. Así, primero debemos verificar la igualdad de las varianzas en el subperiodo, mediante técnicas estadísticas apropiadas. Aunque analizaremos con mayor profundidad este tema más adelante en el capítulo sobre heteroscedasticidad, en el capítulo 8 vimos que la prueba *F* puede utilizarse para este propósito.[20] (Véase el análisis de la prueba de Chow en ese capítulo.) Como demostramos ahí, parece que la varianza del error para los dos periodos no es la misma. En consecuencia, los resultados de la prueba de Chow y la técnica de las variables dicótomas presentadas en este apartado tal vez no sean del todo confiables. Desde luego, el propósito aquí es ilustrar diversas técnicas para resolver el problema (por ejemplo, el problema de la estabilidad estructural). En alguna aplicación particular, tales técnicas quizá no resulten válidas. Pero lo anterior también sucede con la mayoría de las técnicas estadísticas. Por supuesto, hay que tomar las medidas correctivas apropiadas a fin de resolver el problema, como más tarde se hará en el capítulo sobre heteroscedasticidad (no obstante, consulte el ejercicio 9.28).

## Variables dicótomas y autocorrelación

Además de la homoscedasticidad, el modelo de regresión lineal clásico supone que el término de error en los modelos de regresión no está correlacionado. Pero, ¿qué sucede si esto no es así, sobre todo en los modelos que implican regresoras dicótomas? En vista de que analizaremos a profundidad el tema de la autocorrelación en el capítulo respectivo, diferiremos la respuesta hasta ese momento.

## ¿Qué sucede si la variable dependiente es dicótoma?

Hasta ahora hemos considerado modelos en que la regresada es cuantitativa y las regresoras son cuantitativas o cualitativas o de ambos tipos. Pero existen situaciones en que la regresada también puede ser una variable cualitativa o dicótoma. Considere por ejemplo la decisión de un trabajador de participar en la fuerza laboral. La decisión de participar es del tipo sí o no. Será sí, si la persona decide participar, y no, en cualquier otro caso. Por tanto, la variable participación en la fuerza laboral es una variable dicótoma. Desde luego, la decisión de participar en la fuerza laboral depende de diversos factores, como la tasa de salario inicial, la escolaridad y las condiciones del mercado laboral (como las mide la tasa de desempleo).

¿Todavía podemos utilizar los MCO para estimar los modelos de regresión en los que la regresada es dicótoma? Sí, mecánicamente es posible. Pero tales modelos presentan varios problemas estadísticos. Y como hay opciones para la estimación por MCO que no provocan tales inconvenientes, veremos este tema en un capítulo posterior (véase el capítulo 15 sobre los modelos logit y probit). En ese capítulo también estudiaremos los modelos en los cuales la regresada tiene más de dos categorías; por ejemplo, la decisión de ir al trabajo en automóvil, autobús o metro; o la decisión de trabajar tiempo parcial, completo o no trabajar en absoluto. Tales modelos se conocen como modelos con **variable dependiente policótomas**, en contraste con los **modelos con variables dependientes dicótomas**, en los que la variable dependiente tiene sólo dos categorías.

[20] El procedimiento de la prueba de Chow se puede realizar incluso en presencia de heteroscedasticidad, pero entonces se tendrá que utilizar la **prueba de Wald**. Los cálculos matemáticos en los que se basa esta prueba son un tanto complejos. No obstante, en el capítulo sobre heteroscedasticidad volveremos a este tema.

## 9.11 Temas para estudio posterior

En la teoría se analizan diversos temas relativamente avanzados que tienen que ver con las variables dicótomas, como 1) **modelos de parámetros aleatorios** o **variables**, 2) **modelos de regresión cambiantes** y 3) **modelos de desequilibrio**.

En los modelos de regresión considerados en este texto se supone que los parámetros, las $\beta$, son desconocidos pero fijos. Los modelos de coeficientes aleatorios —de los cuales hay diversas versiones— suponen que las $\beta$ también pueden ser aleatorias. El trabajo principal de referencia en esta área es el de Swamy.[21]

En el modelo de variable dicótoma que utiliza interceptos diferenciales al igual que pendientes diferenciales se supone implícitamente que se conoce el punto de ruptura. Por tanto, en el ejemplo de ahorro-ingreso de 1970 a 1995 se dividió el lapso en 1970-1981 y 1982-1995, los periodos anterior y posterior de la recesión, en la creencia de que la recesión de 1982 cambió la relación entre ahorro e ingreso. A veces no es sencillo señalar el momento de la ruptura. La técnica de **modelos de regresión cambiantes** maneja esta situación al permitir que el punto de ruptura sea en sí mismo una variable aleatoria y, mediante un proceso iterativo, determinar cuándo pudo acontecer realmente la ruptura. El trabajo original en esta área se atribuye a Goldfeld y Quandt.[22]

Se requieren técnicas especiales de estimación para lo que se conoce como **situaciones de desequilibrio**, es decir, situaciones en donde los mercados no son claros (la demanda no es igual a la oferta). El ejemplo clásico es el de la demanda y oferta de un *commodity* (bien no diferenciado). La demanda de un *commodity* es función de su precio y de otras variables, y su oferta es también función de su precio y de otras variables, algunas de las cuales son diferentes de las que forman parte de la función de demanda. Ahora, la cantidad en realidad comprada y vendida del *commodity* no necesariamente será igual a la que se obtiene al igualar la demanda a la oferta, lo que genera un desequilibrio. Para un análisis completo de **modelos de desequilibrio**, el lector puede consultar a Quandt.[23]

## 9.12 Ejemplo para concluir

Para concluir este capítulo presentamos un ejemplo que ilustra algunos puntos. La tabla 9.7 proporciona datos sobre una muestra de 261 trabajadores en una ciudad industrial del sur de India en 1990.

Las variables se definen como sigue:

IS = ingreso por salario semanal en rupias
Edad = edad en años
$D_{\text{sexo}}$ = 1 para trabajadores y 0 para trabajadoras
$DE_2$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad hasta primaria
$DE_3$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad hasta secundaria
$DE_4$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad superior al nivel secundaria
$DPT$ = variable dicótoma que toma el valor de 1 para trabajadores con empleo permanente y 0 para eventuales

[21] P.A.V.B. Swamy, *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, Berlín, 1971.

[22] S. Goldfeld y R. Quandt, *Nonlinear Methods in Econometrics*, North Holland, Amsterdam, 1972.

[23] Richard E. Quandt, *The Econometrics of Disequilibrium*, Basil Blackwell, Nueva York, 1988.

**TABLA 9.7** **Indian wage Earners, 1990**

| IS | EDAD | $DE_2$ | $DE_3$ | $DE_4$ | DPT | $D_{SEXO}$ | IS | EDAD | $DE_2$ | $DE_3$ | $DE_4$ | DPT | $D_{SEXO}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 120 | 57 | 0 | 0 | 0 | 0 | 0 | 120 | 21 | 0 | 0 | 0 | 0 | 0 |
| 224 | 48 | 0 | 0 | 1 | 1 | 0 | 25 | 18 | 0 | 0 | 0 | 0 | 1 |
| 132 | 38 | 0 | 0 | 0 | 0 | 0 | 25 | 11 | 0 | 0 | 0 | 0 | 1 |
| 75 | 27 | 0 | 1 | 0 | 0 | 0 | 30 | 38 | 0 | 0 | 0 | 1 | 1 |
| 111 | 23 | 0 | 1 | 0 | 0 | 1 | 30 | 17 | 0 | 0 | 0 | 1 | 1 |
| 127 | 22 | 0 | 1 | 0 | 0 | 0 | 122 | 20 | 0 | 0 | 0 | 0 | 0 |
| 30 | 18 | 0 | 0 | 0 | 0 | 0 | 288 | 50 | 0 | 1 | 0 | 1 | 0 |
| 24 | 12 | 0 | 0 | 0 | 0 | 0 | 75 | 45 | 0 | 0 | 0 | 0 | 1 |
| 119 | 38 | 0 | 0 | 0 | 1 | 0 | 79 | 60 | 0 | 0 | 0 | 0 | 0 |
| 75 | 55 | 0 | 0 | 0 | 0 | 0 | 85.3 | 26 | 1 | 0 | 0 | 0 | 1 |
| 324 | 26 | 0 | 1 | 0 | 0 | 0 | 350 | 42 | 0 | 1 | 0 | 1 | 0 |
| 42 | 18 | 0 | 0 | 0 | 0 | 0 | 54 | 62 | 0 | 0 | 0 | 1 | 0 |
| 100 | 32 | 0 | 0 | 0 | 0 | 0 | 110 | 23 | 0 | 0 | 0 | 0 | 0 |
| 136 | 41 | 0 | 0 | 0 | 0 | 0 | 342 | 56 | 0 | 0 | 0 | 1 | 0 |
| 107 | 48 | 0 | 0 | 0 | 0 | 0 | 77.5 | 19 | 0 | 0 | 0 | 1 | 0 |
| 50 | 16 | 1 | 0 | 0 | 0 | 1 | 370 | 46 | 0 | 0 | 0 | 0 | 0 |
| 90 | 45 | 0 | 0 | 0 | 0 | 0 | 156 | 26 | 0 | 0 | 0 | 1 | 0 |
| 377 | 46 | 0 | 0 | 0 | 1 | 0 | 261 | 23 | 0 | 0 | 0 | 0 | 0 |
| 150 | 30 | 0 | 1 | 0 | 0 | 0 | 54 | 16 | 0 | 1 | 0 | 0 | 0 |
| 162 | 40 | 0 | 0 | 0 | 0 | 0 | 130 | 33 | 0 | 0 | 0 | 0 | 0 |
| 18 | 19 | 1 | 0 | 0 | 0 | 0 | 112 | 27 | 1 | 0 | 0 | 0 | 0 |
| 128 | 25 | 1 | 0 | 0 | 0 | 0 | 82 | 22 | 1 | 0 | 0 | 0 | 0 |
| 47.5 | 46 | 0 | 0 | 0 | 0 | 1 | 385 | 30 | 0 | 1 | 0 | 1 | 0 |
| 135 | 25 | 0 | 1 | 0 | 0 | 0 | 94.3 | 22 | 0 | 0 | 1 | 1 | 1 |
| 400 | 57 | 0 | 0 | 0 | 1 | 0 | 350 | 57 | 0 | 0 | 0 | 1 | 0 |
| 91.8 | 35 | 0 | 0 | 1 | 1 | 0 | 108 | 26 | 0 | 0 | 0 | 0 | 0 |
| 140 | 44 | 0 | 0 | 0 | 1 | 0 | 20 | 14 | 0 | 0 | 0 | 0 | 0 |
| 49.2 | 22 | 0 | 0 | 0 | 0 | 0 | 53.8 | 14 | 0 | 0 | 0 | 0 | 1 |
| 30 | 19 | 1 | 0 | 0 | 0 | 0 | 427 | 55 | 0 | 0 | 0 | 1 | 0 |
| 40.5 | 37 | 0 | 0 | 0 | 0 | 1 | 18 | 12 | 0 | 0 | 0 | 0 | 0 |
| 81 | 20 | 0 | 0 | 0 | 0 | 0 | 120 | 38 | 0 | 0 | 0 | 0 | 0 |
| 105 | 40 | 0 | 0 | 0 | 0 | 0 | 40.5 | 17 | 0 | 0 | 0 | 0 | 0 |
| 200 | 30 | 0 | 0 | 0 | 0 | 0 | 375 | 42 | 1 | 0 | 0 | 1 | 0 |
| 140 | 30 | 0 | 0 | 0 | 1 | 0 | 120 | 34 | 0 | 0 | 0 | 0 | 0 |
| 80 | 26 | 0 | 0 | 0 | 0 | 0 | 175 | 33 | 1 | 0 | 0 | 1 | 0 |
| 47 | 41 | 0 | 0 | 0 | 0 | 1 | 50 | 26 | 0 | 0 | 0 | 0 | 1 |
| 125 | 22 | 0 | 0 | 0 | 0 | 0 | 100 | 33 | 1 | 0 | 0 | 1 | 0 |
| 500 | 21 | 0 | 0 | 0 | 0 | 0 | 25 | 22 | 0 | 0 | 0 | 1 | 1 |
| 100 | 19 | 0 | 0 | 0 | 0 | 0 | 40 | 15 | 0 | 0 | 0 | 1 | 0 |
| 105 | 35 | 0 | 0 | 0 | 0 | 0 | 65 | 14 | 0 | 0 | 0 | 1 | 0 |
| 300 | 35 | 0 | 1 | 0 | 1 | 0 | 47.5 | 25 | 0 | 0 | 0 | 1 | 1 |
| 115 | 33 | 0 | 1 | 0 | 1 | 1 | 163 | 25 | 0 | 0 | 0 | 1 | 0 |
| 103 | 27 | 0 | 0 | 1 | 1 | 1 | 175 | 50 | 0 | 0 | 0 | 1 | 1 |
| 190 | 62 | 1 | 0 | 0 | 0 | 0 | 150 | 24 | 0 | 0 | 0 | 1 | 1 |
| 62.5 | 18 | 0 | 1 | 0 | 0 | 0 | 163 | 28 | 0 | 0 | 0 | 1 | 0 |
| 50 | 25 | 1 | 0 | 0 | 0 | 0 | 163 | 30 | 1 | 0 | 0 | 1 | 0 |
| 273 | 43 | 0 | 0 | 1 | 1 | 1 | 50 | 25 | 0 | 0 | 0 | 1 | 1 |
| 175 | 40 | 0 | 1 | 0 | 1 | 0 | 395 | 45 | 0 | 1 | 0 | 1 | 0 |
| 117 | 26 | 1 | 0 | 0 | 1 | 0 | 175 | 40 | 0 | 0 | 0 | 1 | 1 |
| 950 | 47 | 0 | 0 | 1 | 0 | 0 | 87.5 | 25 | 1 | 0 | 0 | 0 | 0 |
| 100 | 30 | 0 | 0 | 0 | 0 | 0 | 75 | 18 | 0 | 0 | 0 | 0 | 0 |
| 140 | 30 | 0 | 0 | 0 | 0 | 0 | 163 | 24 | 0 | 0 | 0 | 1 | 0 |
| 97 | 25 | 0 | 1 | 0 | 0 | 0 | 325 | 55 | 0 | 0 | 0 | 1 | 0 |
| 150 | 36 | 0 | 0 | 0 | 0 | 0 | 121 | 27 | 0 | 1 | 0 | 0 | 0 |
| 25 | 28 | 0 | 0 | 0 | 0 | 1 | 600 | 35 | 1 | 0 | 0 | 0 | 0 |
| 15 | 13 | 0 | 0 | 0 | 0 | 1 | 52 | 19 | 0 | 0 | 0 | 0 | 0 |
| 131 | 55 | 0 | 0 | 0 | 0 | 0 | 117 | 28 | 1 | 0 | 0 | 0 | 0 |

La categoría de referencia es trabajadores sin educación primaria y con empleo eventual. El interés se centra en la relación entre salario semanal y edad, sexo, nivel de escolaridad y titularidad de empleo. Con este fin, estimamos el siguiente modelo de regresión:

$$\ln \text{IS}_i = \beta_1 + \beta_2 \text{EDAD}_i + \beta 3 \text{D}_{\text{sexo}} + \beta_4 DE_2 + \beta_5 DE_3 + \beta_6 DE_4 + \beta_7 DPT + u_i$$

En la línea de la bibliografía de economía laboral, el logaritmo (natural) de los salarios se expresa como función de las variables explicativas. Como señalamos en el capítulo 6, la distribución del tamaño de las variables, como los salarios, tiende a ser asimétrica; las transformaciones logarítmicas de dichas variables reducen tanto la asimetría como la heteroscedasticidad.

Con *EViews6* obtenemos los siguientes resultados de regresión:

```
Variable dependiente: Ln(IS)
Método: Mínimos cuadrados
Muestra: 1 261
Observaciones incluidas: 261
```

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | 3.706872 | 0.113845 | 32.56055 | 0.0000 |
| EDAD | 0.026549 | 0.003117 | 8.516848 | 0.0000 |
| $D_{\text{SEXO}}$ | -0.656338 | 0.088796 | -7.391529 | 0.0000 |
| $DE_2$ | 0.113862 | 0.098542 | 1.155473 | 0.2490 |
| $DE_3$ | 0.412589 | 0.096383 | 4.280732 | 0.0000 |
| $DE_4$ | 0.554129 | 0.155224 | 3.569862 | 0.0004 |
| *DPT* | 0.558348 | 0.079990 | 6.980248 | 0.0000 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.534969 | Media de la variable dependiente | 4.793390 |
| R cuadrada ajustada | 0.523984 | Desviación estándar de la variable dependiente | 0.834277 |
| Error estándar de la regresión | 0.575600 | | |
| Suma de cuadrados residual | 84.15421 | Criterio de información de Akaike | 1.759648 |
| Log verosimilitud | -222.6340 | Criterio de Schwarz | 1.855248 |
| Estadístico *F* | 48.70008 | Criterio de Hannan-Quinn | 1.798076 |
| Probabilidad (estadístico *F*) | 0.000000 | Estadístico de Durbin-Watson | 1.853361 |

Estos resultados muestran que el logaritmo de los salarios se relaciona positivamente con la edad, nivel de escolaridad y permanencia en el empleo, pero negativamente con el sexo, lo que no es sorprendente. Aunque al parecer no existe diferencia práctica entre los salarios semanales de los trabajadores con educación primaria completa y sin terminar, los de los trabajadores con educación secundaria son más altos, y mucho más los de los trabajadores con educación superior.

Los coeficientes de las variables dicótomas deben interpretarse como valores diferenciales de la categoría de referencia. Así, el coeficiente de la variable *DPT* indica que los trabajadores con empleo permanente ganan, en promedio, más dinero que los trabajadores eventuales.

Como vimos en el capítulo 6, en un modelo log-lineal (la variable dependiente en la forma logarítmica y las variables explicativas en la forma lineal), el coeficiente de la pendiente de una

variable explicativa representa semielasticidad, es decir, da el cambio relativo o porcentual en la variable dependiente por cada cambio unitario en el valor de la variable explicativa. Sin embargo, como se anotó en el texto, cuando la variable explicativa es dicótoma, hay que tener mucho cuidado. Aquí se debe tomar el antilogaritmo del coeficiente estimado de la variable dicótoma, restarle 1 y multiplicar el resultado por 100. Por consiguiente, para averiguar el cambio porcentual en el salario semanal de los trabajadores con empleo permanente en comparación con los eventuales, tomamos el antilogaritmo del coeficiente de *DPT* 0.558348, restamos 1 y luego multiplicamos la diferencia por 100. En el ejemplo en cuestión, esto resulta $(e^{0.558348} - 1) = (1.74778 - 1) = 0.74778$, o casi 75%. Se recomienda al lector calcular estos cambios porcentuales para las demás variables dicótomas del modelo.

Los resultados muestran que el sexo y el nivel de escolaridad tienen efectos diferenciales en las percepciones semanales. ¿Es posible una interacción entre sexo y nivel de escolaridad? ¿Los trabajadores con nivel de escolaridad alto ganan mejores salarios semanales que las trabajadoras con el mismo nivel de escolaridad? Para examinar esta posibilidad extendemos la anterior regresión de salarios para añadir la interacción entre sexo y nivel de escolaridad. Los resultados de la regresión son los siguientes:

```
Variable dependiente: Ln(IS)
Método: Mínimos cuadrados
Muestra: 1 261
Observaciones incluidas: 261
```

| | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| C | 3.717540 | 0.114536 | 32.45734 | 0.0000 |
| EDAD | 0.027051 | 0.003133 | 8.634553 | 0.0000 |
| $D_{SEXO}$ | -0.758975 | 0.110410 | -6.874148 | 0.0000 |
| $DE_2$ | 0.088923 | 0.106827 | 0.832402 | 0.4060 |
| $DE_3$ | 0.350574 | 0.104309 | 3.360913 | 0.0009 |
| $DE_4$ | 0.438673 | 0.186996 | 2.345898 | 0.0198 |
| $D_{SEXO}*DE_2$ | 0.114908 | 0.275039 | 0.417788 | 0.6765 |
| $D_{SEXO}*DE_3$ | 0.391052 | 0.259261 | 1.508337 | 0.1327 |
| $D_{SEXO}*DE_4$ | 0.369520 | 0.313503 | 1.178681 | 0.2396 |
| $DPT$ | 0.551658 | 0.080076 | 6.889198 | 0.0000 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.540810 | Media de la variable dependiente | 4.793390 |
| R cuadrada ajustada | 0.524345 | Desviación estándar de la variable dependiente | 0.834277 |
| Error estándar de la regresión | 0.575382 | Criterio de información de Akaike | 1.769997 |
| Suma de cuadrados residual | 83.09731 | Criterio de Schwarz | 1.906569 |
| Log verosimilitud | -220.9847 | Criterio de Hannan-Quinn | 1.824895 |
| Estadístico $F$ | 32.84603 | Estadístico de Durbin-Watson | 1.856588 |
| Probabilidad (estadístico $F$) | 0.000000 | | |

Aunque las variables dicótomas de interacción revelan cierta interacción entre sexo y nivel de escolaridad, el efecto no es estadísticamente significativo, pues ningún coeficiente de interacción es estadísticamente significativo por sí solo.

Es interesante notar que, si eliminamos las variables dicótomas de escolaridad pero conservamos las de interacción, obtenemos los siguientes resultados:

```
Variable dependiente: LOG(IS)
Método: Mínimos cuadrados
Muestra: 1 261
Observaciones incluidas: 261
```

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | 3.836483 | 0.106785 | 35.92725 | 0.0000 |
| EDAD | 0.025990 | 0.003170 | 8.197991 | 0.0000 |
| $D_{\text{SEXO}}$ | -0.868617 | 0.106429 | -8.161508 | 0.0000 |
| $D_{\text{SEXO}}*DE_2$ | 0.200823 | 0.259511 | 0.773851 | 0.4397 |
| $D_{\text{SEXO}}*DE_3$ | 0.716722 | 0.245021 | 2.925140 | 0.0038 |
| $D_{\text{SEXO}}*DE_4$ | 0.752652 | 0.265975 | 2.829789 | 0.0050 |
| *DPT* | 0.627272 | 0.078869 | 7.953332 | 0.0000 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.514449 | Media de la variable dependiente | 4.793390 |
| R cuadrada ajustada | 0.502979 | Desviación estándar de la variable dependiente | 0.834277 |
| Error estándar de la regresión | 0.488163 | Criterio de información de Akaike | 1.802828 |
| Suma de cuadrados residual | 87.86766 | Criterio de Schwarz | 1.898429 |
| Log verosimilitud | -228.2691 | Criterio de Hannan-Quinn | 1.841257 |
| Estadístico *F* | 44.85284 | Estadístico de Durbin-Watson | 1.873421 |
| Probabilidad (estadístico *F*) | 0.000000 | | |

Ahora parece que las variables dicótomas de escolaridad, por sí solas, no tienen efecto en el salario semanal, pero introducidas en formato interactivo sí producen efectos. Como muestra este ejercicio, hay que tener cuidado con las variables dicótomas. Como ejercicio, el lector deberá averiguar si las variables dicótomas de escolaridad interactúan con DPT.

## Resumen y conclusiones

1. Las variables dicótomas con valores de 1 y 0 (o sus transformaciones lineales) son un medio de introducir regresoras cualitativas en el análisis de regresión.
2. Las variables dicótomas son un mecanismo de clasificación de datos, pues permiten dividir una muestra en diversos subgrupos con base en cualidades o atributos (sexo, estado civil, raza, religión, etc.) e *implícitamente* permiten efectuar regresiones individuales para cada subgrupo. Si hay diferencias en la respuesta de la variable regresada a la variación en las variables cuantitativas en los diversos subgrupos, éstas se reflejarán en las diferencias en los interceptos o en los coeficientes de las pendientes, o en ambos, de las regresiones de los diversos subgrupos.
3. Aunque es una herramienta versátil, la técnica de variable dicótoma debe manejarse con cuidado. *Primero*, si la regresión contiene un término constante, el número de variables dicótomas debe ser menor que el número de clasificaciones de cada variable cualitativa. *Segundo*, el coeficiente que acompaña las variables dicótomas *siempre* debe interpretarse en relación con el grupo base o de referencia, es decir, con el grupo que adquiere el valor de cero. La base elegida depende del propósito de la investigación en curso. *Finalmente*, si un modelo tiene diversas variables cualitativas con diversas categorías, la introducción de las variables dicótomas puede consumir un gran número de grados de libertad. Por consiguiente, siempre se debe ponderar el número de variables dicótomas que se van a introducir respecto del número total de observaciones disponible para el análisis.

4. En este capítulo se consideraron sólo algunas de las diversas aplicaciones de la técnica de variables dicótomas: 1) comparación de dos (o más) regresiones, 2) desestacionalización de datos de series de tiempo, 3) variables dicótomas interactivas, 4) interpretación de las variables dicótomas en los modelos semilogarítmicos y 5) modelos de regresión lineal por segmentos.
5. También recomendamos mucho tener precaución al utilizar las variables dicótomas en situaciones de heteroscedasticidad y autocorrelación. Pero, como estudiaremos estos temas con mucho detalle en capítulos subsecuentes, abordaremos esto a su debido tiempo.

## EJERCICIOS

### Preguntas

9.1. Si cuenta con datos mensuales de distintos años, ¿cuántas variables dicótomas introducirá para probar las siguientes hipótesis?:

*a*) Los 12 meses del año presentan patrones estacionales.

*b*) Sólo febrero, abril, junio, agosto, octubre y diciembre presentan patrones estacionales.

9.2. Considere los siguientes resultados de regresión (las razones *t* están entre paréntesis):*

$$\begin{aligned}
\hat{Y}_i &= 1\,286 \quad + 104.97X_{2i} - \quad 0.026X_{3i} + \quad 1.20X_{4i} + \quad 0.69X_{5i} \\
t &= \;(4.67) \qquad (3.70) \qquad (-3.80) \qquad (0.24) \qquad (0.08) \\
&\quad -19.47X_{6i} + 266.06X_{7i} \qquad - 118.64X_{8i} - 110.61X_{9i} \\
&\quad (-0.40) \qquad (6.94) \qquad\qquad (-3.04) \qquad (-6.14) \\
&\qquad\qquad R^2 = 0.383 \qquad n = 1\,543
\end{aligned}$$

donde $Y$ = horas de trabajo anuales deseadas por la esposa, calculadas como horas normales de trabajo al año, más las semanas invertidas en buscar trabajo

$X_2$ = ingresos promedio reales por hora después de impuestos de la esposa

$X_3$ = ingresos anuales reales después de impuestos del esposo en el año anterior

$X_4$ = edad de la esposa en años

$X_5$ = años completos de escolaridad de la esposa

$X_6$ = variable de actitud; 1 si la entrevistada considera correcto que una mujer trabaje si así lo desea y su esposo está de acuerdo; 0 en cualquier otro caso

$X_7$ = variable de actitud; 1 si el esposo de la entrevistada favoreció que su esposa trabajara; 0 en cualquier otro caso

$X_8$ = número de hijos menores de 6 años

$X_9$ = número de hijos de 6 a 13 años

*a*) ¿Los signos de los coeficientes de las diversas regresoras no dicótomas tienen algún sentido económico? Justifique su respuesta.

*b*) ¿Cómo interpretaría las variables dicótomas $X_6$ y $X_7$? ¿Son estadísticamente significativas? Como la muestra es muy grande, tal vez se utilice la regla práctica "2 *t*" para responder la última pregunta.

*c*) ¿Por qué cree que, en este estudio, las variables edad y escolaridad no son factores significativos en la decisión de la mujer de participar en la fuerza laboral?

---

* Jane Leuthold, "The Effect of Taxation on the Hours Worked by Married Women", *Industrial Labor Relations Review*, núm. 4, julio de 1978, pp. 520-526 (la notación se cambió para adaptarla al formato del libro).

**TABLA 9.8**
**Matriz de datos para la regresión del ejercicio 9.3**

Fuente: Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958-1971", *The Economic Journal*, vol. 82, marzo de 1972, p. 202.

| Año y trimestre | Tasa de desempleo TD, % | Tasa de puestos vacantes V, % | *D* | *DV* | Año y trimestre | Tasa de desempleo TD, % | Tasa de puestos vacantes V, % | *D* | *DV* |
|---|---|---|---|---|---|---|---|---|---|
| 1958–IV | 1.915 | 0.510 | 0 | 0 | 1965–I | 1.201 | 0.997 | 0 | 0 |
| 1959–I | 1.876 | 0.541 | 0 | 0 | –II | 1.192 | 1.035 | 0 | 0 |
| –II | 1.842 | 0.541 | 0 | 0 | –III | 1.259 | 1.040 | 0 | 0 |
| –III | 1.750 | 0.690 | 0 | 0 | –IV | 1.192 | 1.086 | 0 | 0 |
| –IV | 1.648 | 0.771 | 0 | 0 | 1966–I | 1.089 | 1.101 | 0 | 0 |
| 1960–I | 1.450 | 0.836 | 0 | 0 | –II | 1.101 | 1.058 | 0 | 0 |
| –II | 1.393 | 0.908 | 0 | 0 | –III | 1.243 | 0.987 | 0 | 0 |
| –III | 1.322 | 0.968 | 0 | 0 | –IV | 1.623 | 0.819 | 1 | 0.819 |
| –IV | 1.260 | 0.998 | 0 | 0 | 1967–I | 1.821 | 0.740 | 1 | 0.740 |
| 1961–I | 1.171 | 0.968 | 0 | 0 | –II | 1.990 | 0.661 | 1 | 0.661 |
| –II | 1.182 | 0.964 | 0 | 0 | –III | 2.114 | 0.660 | 1 | 0.660 |
| –III | 1.221 | 0.952 | 0 | 0 | –IV | 2.115 | 0.698 | 1 | 0.698 |
| –IV | 1.340 | 0.849 | 0 | 0 | 1968–I | 2.150 | 0.695 | 1 | 0.695 |
| 1962–I | 1.411 | 0.748 | 0 | 0 | –II | 2.141 | 0.732 | 1 | 0.732 |
| –II | 1.600 | 0.658 | 0 | 0 | –III | 2.167 | 0.749 | 1 | 0.749 |
| –III | 1.780 | 0.562 | 0 | 0 | –IV | 2.107 | 0.800 | 1 | 0.800 |
| –IV | 1.941 | 0.510 | 0 | 0 | 1969–I | 2.104 | 0.783 | 1 | 0.783 |
| 1963–I | 2.178 | 0.510 | 0 | 0 | –II | 2.056 | 0.800 | 1 | 0.800 |
| –II | 2.067 | 0.544 | 0 | 0 | –III | 2.170 | 0.794 | 1 | 0.794 |
| –III | 1.942 | 0.568 | 0 | 0 | –IV | 2.161 | 0.790 | 1 | 0.790 |
| –IV | 1.764 | 0.677 | 0 | 0 | 1970–I | 2.225 | 0.757 | 1 | 0.757 |
| 1964–I | 1.532 | 0.794 | 0 | 0 | –II | 2.241 | 0.746 | 1 | 0.746 |
| –II | 1.455 | 0.838 | 0 | 0 | –III | 2.366 | 0.739 | 1 | 0.739 |
| –III | 1.409 | 0.885 | 0 | 0 | –IV | 2.324 | 0.707 | 1 | 0.707 |
| –IV | 1.296 | 0.978 | 0 | 0 | 1971–I | 2.516* | 0.583* | 1 | 0.583* |
| | | | | | –II | 2.909* | 0.524* | 1 | 0.524* |

*Estimados preliminares.

9.3. Considere los siguientes resultados de una regresión.* (Los datos reales se proporcionan en la tabla 9.8.)

$$\begin{aligned} \widehat{\text{TD}}_t = {} & 2.7491 + 1.1507D_t - 1.5294V_t - 0.8511(D_tV_t) \\ t = {} & (26.896) \quad (3.6288) \quad (-12.5552) \quad (-1.9819) \\ & R^2 = 0.9128 \end{aligned}$$

donde TD = tasa de desempleo, %
$V$ = tasa de puestos vacantes, %
$D$ = 1, para el periodo que comienza el cuarto trimestre de 1966
= 0, para el periodo anterior al cuatro trimestre de 1966
$t$ = tiempo, medido en trimestres

*Nota:* En el cuarto trimestre de 1966, el entonces gobierno laborista liberalizó la Ley de Seguro Nacional: reemplazó el sistema de tasa fija para prestaciones de desempleo de corto plazo por un sistema mixto de tasa fija y prestaciones relacionadas con los ingresos (anteriores), el cual incrementó el nivel de las prestaciones de desempleo.

* Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958-1971", *The Economic Journal*, vol. 82, marzo de 1972, pp. 195-202.

*a*) ¿Cuáles son las expectativas *a priori* respecto de la relación entre las tasas de desempleo y de vacantes?

*b*) Si la tasa de vacantes se mantiene constante, ¿cuál es la tasa promedio de desempleo para el periodo que comienza el cuarto trimestre de 1966? ¿Es estadísticamente distinto del periodo anterior al cuarto trimestre de 1966? ¿Cómo sabe?

*c*) ¿Las pendientes para el periodo anterior y posterior al cuarto trimestre de 1966 son estadísticamente distintas? ¿Cómo sabe?

*d*) ¿Se puede concluir con toda seguridad, a partir de este estudio, que los generosos beneficios del desempleo propician tasas más altas de vacantes? ¿Lo anterior tiene algún sentido económico?

9.4. Con base en datos anuales para el periodo 1972-1979, William Nordhaus estimó el siguiente modelo para explicar el comportamiento del precio del petróleo de la OPEP (errores estándar entre paréntesis):*

$$\begin{aligned} \hat{y}_t &= \ 0.3x_{1t} + \ 5.22x_{2t} \\ \text{ee} &= (0.03) \quad\ (0.50) \end{aligned}$$

donde $y$ = diferencia entre el precio del año en curso y el año anterior (dólares por barril)
$x_1$ = diferencia entre el precio para entrega inmediata del año en curso y el precio de la OPEP en el año anterior
$x_2$ = 1 para el año 1974 y 0 de lo contrario

Interprete este resultado y muestre el resultado gráficamente. ¿Qué indican estos resultados acerca del poder monopólico de la OPEP?

9.5. Considere el siguiente modelo

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + u_i$$

donde $Y$ = salario anual de un profesor universitario
$X$ = años de experiencia docente
$D$ = variable dicótoma para el sexo

Considere tres formas de definir la variable dicótoma:

*a*) $D = 1$ si es hombre; 0 si es mujer.

*b*) $D = 1$ si es mujer; 0 si es hombre.

*c*) $D = 1$ si es mujer; $-1$ si es hombre.

Interprete el modelo de regresión anterior para cada asignación de variable dicótoma. ¿Es preferible un método en vez de otro? Justifique su respuesta.

9.6. Consulte la regresión (9.7.3). ¿Cómo probaría la hipótesis de que los coeficientes de $D_2$ y $D_3$ son iguales? ¿Y que los coeficientes $D_2$ y $D_4$ son iguales? Si el coeficiente de $D_3$ es estadísticamente distinto del de $D_2$, y el coeficiente de $D_4$ es diferente del de $D_2$, ¿significa que los coeficientes $D_3$ y $D_4$ son también distintos?
*Sugerencia:* $\text{var}(A \pm B) = \text{var}(A) + \text{var}(B) \pm 2\,\text{cov}(A, B)$

9.7. Consulte el ejemplo de ahorro-ingreso en Estados Unidos analizado en la sección 9.5.

*a*) ¿Cómo obtendría los errores estándar de los coeficientes de regresión dados en (9.5.5) y (9.5.6), los cuales se obtuvieron de la regresión agrupada (9.5.4)?

*b*) Para obtener respuestas numéricas, ¿qué información adicional se requeriría, en caso de necesitarse?

* "Oil and Economic Performance in Industrial Countries", *Brookings Papers on Economic Activity*, 1980, pp. 341-388.

9.8. En su estudio sobre las horas de trabajo dedicadas por el FDIC (Federal Deposit Insurance Corporation) al análisis de 91 bancos, R.J. Miller estimó la siguiente función:*

$$
\begin{aligned}
\widehat{\ln Y} = 2.41 &+ \underset{(0.0477)}{0.3674 \ln X_1} + \underset{(0.0628)}{0.2217 \ln X_2} + \underset{(0.0287)}{0.0803 \ln X_3} \\
&\underset{(0.2905)}{-0.1755 D_1} + \underset{(0.1044)}{0.2799 D_2} + \underset{(0.1657)}{0.5634 D_3} - \underset{(0.0787)}{0.2572 D_4} \\
&R^2 = 0.766
\end{aligned}
$$

donde $Y$ = horas-hombre del examinador del FDIC
$X_1$ = activos totales del banco
$X_2$ = número total de oficinas del banco
$X_3$ = razón de préstamos clasificados a préstamos totales del banco
$D_1$ = 1 si la administración se calificó "muy buena"
$D_2$ = 1 si la administración se calificó "buena"
$D_3$ = 1 si la administración se calificó "satisfactoria"
$D_4$ = 1 si la evaluación se realizó junto con el estado

Las cifras en paréntesis son los errores estándar estimados.

*a*) Interprete estos resultados.

*b*) ¿Hay algún problema en la interpretación de las variables dicótomas en este modelo por estar $Y$ en forma logarítmica?

*c*) ¿Cómo interpretaría los coeficientes de las variables dicótomas?

9.9. Para evaluar el efecto de la política del gobierno federal sobre la liberalización de tasas de interés iniciada en julio de 1979, Sidney Langer, alumna del autor, estimó el siguiente modelo para el periodo trimestral comprendido entre 1975-III y 1983-II.†

$$
\begin{aligned}
\hat{Y}_t = \;& 8.5871 - 0.1328 P_t - 0.7102 \text{Un}_t - 0.2389 M_t \\
\text{ee} = \;& (1.9563) \quad (0.0992) \quad (0.1909) \quad (0.0727) \\
&+ 0.6592 Y_{t-1} + 2.5831 \text{Dic}_t \quad R^2 = 0.9156 \\
&\quad (0.1036) \quad (0.7549)
\end{aligned}
$$

donde $Y$ = tasa de bonos del Tesoro a tres meses
$P$ = tasa de inflación esperada
Un = tasa de desempleo ajustada por estacionalidad
$M$ = cambios en la base monetaria
Dic = variable dicótoma, que adquiere el valor de 1 para las observaciones que empiezan el 1 de julio de 1979

*a*) Interprete estos resultados.

*b*) ¿Cuál fue el efecto de la liberación de la tasa? ¿Tienen sentido económico los resultados?

*c*) Los coeficientes de $P_t$, $\text{Un}_t$ y $M_t$ son negativos. ¿Puede ofrecer un razonamiento económico?

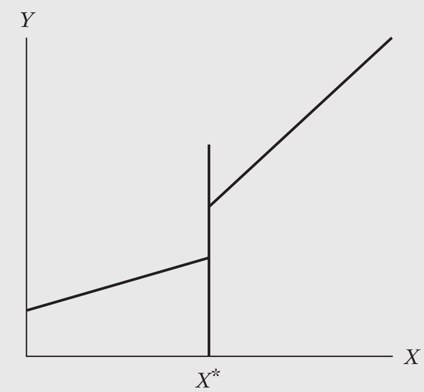
9.10. Consulte la regresión por segmentos analizada en el texto. Suponga que no sólo hay un cambio en el coeficiente de la pendiente en $X^*$, sino también un salto en la línea de regresión, como se muestra en la figura 9.7. ¿Cómo modificaría (9.8.1) para considerar el salto en la línea de regresión en $X^*$?

---

* "Examination of Man-Hour Cost for Independent, Joint, and Divided Examination Programs", *Journal of Bank Research*, vol. 11, 1980, pp. 28-35. *Nota:* Las notaciones se modificaron para adaptarlas a la notación del texto.

† Sidney Langer, "Interest Rate Deregulation and Short-Term Interest Rates", trabajo universitario inédito.

**FIGURA 9.7**
Regresión lineal por segmentos discontinua.




9.11. *Determinantes del precio por onza de cola*. Cathy Schaefer, alumna del autor, estimó la siguiente regresión con base en datos transversales de 77 observaciones:*

$$P_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \mu_i$$

donde $P_i$ = precio por onza de cola
$D_{1i}$ = 001 si es almacén de descuento
= 010 si es almacén de cadena
= 100 si es tienda
$D_{2i}$ = 10 si es un producto de marca
= 01 si es un producto sin marca
$D_{3i}$ = 0001 botella de 67.6 onzas (2 litros)
= 0010 botellas de 28-33.8 onzas (*Nota:* 33.8 oz = 1 litro)
= 0100 botellas de 16 onzas
= 1 000 latas de 12 onzas

Los resultados fueron los siguientes:

$$\hat{P}_i = 0.0143 - 0.000004D_{1i} + 0.0090D_{2i} + 0.00001D_{3i}$$

| | | | |
|---|---|---|---|
| ee = | (0.00001) | (0.00011) | (0.00000) |
| $t$ = | (−0.3837) | (8.3927) | (5.8125) |

$R^2 = 0.6033$

*Nota:* Los errores estándar se muestran sólo con cinco decimales.

*a*) Comente sobre la forma de introducir las variables dicótomas en el modelo.

*b*) Si suponemos que el procedimiento de variables dicótomas es aceptable, ¿cómo interpreta los resultados?

*c*) El coeficiente de $D_3$ es positivo y estadísticamente significativo. ¿Cómo interpreta este resultado?

9.12. Con base en los datos de 101 países sobre el ingreso *per cápita* en dólares ($X$) y la esperanza de vida en años ($Y$) a principios de la década de 1970, Sen y Srivastava obtuvieron los siguientes resultados de regresión:†

$$\hat{Y}_i = -2.40 + 9.39 \ln X_i - 3.36\, [D_i(\ln X_i - 7)]$$

$$\text{ee} = (4.73) \quad (0.859) \quad (2.42) \qquad R^2 = 0.752$$

donde $D_i = 1$ si $\ln X_i > 7$, y $D_i = 0$ de lo contrario. *Nota:* Cuando $\ln X_i = 7$, $X = \$1\,097$ (aproximadamente).

* Cathy Schaefer, "Price Per Ounce of Cola Beverage as a Function of Place of Purchase, Size of Container, and Branded or Unbranded Product", trabajo universitario inédito.

† Ashish Sen y Muni Srivastava, *Regression Analysis: Theory, Methods, and Applications*, Springer-Verlag, Nueva York, 1990, p. 92. Se cambió la notación.

*a*) ¿Cuál(es) puede(*n*) ser la(s) razón(razones) para introducir la variable ingreso en forma logarítmica?

*b*) ¿Cómo interpretaría el coeficiente 9.39 del ln $X_i$?

*c*) ¿Cuál puede ser la razón para introducir la regresora $D_i(\ln X_i - 7)$? ¿Cómo explica esta regresora? ¿Y cómo interpreta el coeficiente −3.36 de esta regresora? (*Sugerencia:* regresión lineal por tramos.)

*d*) Si suponemos un ingreso *per cápita* de \$1 097 como línea divisoria entre los países más pobres y los más ricos, ¿cómo derivaría la regresión para países cuyo ingreso *per cápita* es menor que \$1 097 y la regresión para países cuyo ingreso *per cápita* es mayor que \$1 097?

*e*) ¿Qué conclusiones generales obtiene del resultado de la regresión de este problema?

9.13. Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

donde $D_i = 0$ para las primeras 20 observaciones y $D_i = 1$ para las 30 observaciones restantes. También se le indica que la var $(u_i^2) = 300$.

*a*) ¿Cómo interpretaría $\beta_1$ y $\beta_2$?

*b*) ¿Cuáles son los valores promedio de los dos grupos?

*c*) ¿Cómo calcularía la varianza de $(\hat{\beta}_1 + \hat{\beta}_2)$? *Nota:* Un dato que se le da es que la cov $(\hat{\beta}_1, \hat{\beta}_2) = -15$.

9.14. Para evaluar el efecto de las leyes estatales de derecho al trabajo referentes a los sindicatos (las cuales no obligan a pertenecer a un sindicato como condición para el empleo), se obtuvieron los siguientes resultados de regresión, a partir de datos de 50 estados de Estados Unidos durante 1982:*

$$\begin{aligned}\widehat{\text{PSP}}_i &= 19.8066 - 9.3917\ \text{DAT}_i \\ t &= (17.0352) \quad (-5.1086) \\ & \qquad\qquad\qquad r^2 = 0.3522\end{aligned}$$

donde PSP = porcentaje de los empleados en el sector privado afiliados a sindicatos en 1982, y DAT = 1 si hay leyes de derecho al trabajo y 0 en otro caso. *Nota:* En 1982, dichas leyes estaban vigentes en 20 estados de ese país.

*a*) *A priori*, ¿cuál es la relación esperada entre PSP y DAT?

*b*) ¿Los resultados de la regresión apoyan las expectativas anteriores?

*c*) Interprete los resultados de la regresión.

*d*) ¿Cuál es el porcentaje promedio de los empleados del sector privado pertenecientes a sindicatos, en los estados en donde no están vigentes las leyes sobre el derecho al trabajo?

9.15. En el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

*Y* representa el salario por hora en dólares, y *D* es la variable dicótoma, que toma el valor de 1 si es un titulado universitario y 0 si es un titulado de escuela media superior. Con las fórmulas de MCO del capítulo 3, demuestre que $\hat{\beta}_1 = \bar{Y}_{\text{hg}}$ y $\hat{\beta}_2 = \bar{Y}_{\text{cg}} - \bar{Y}_{\text{hg}}$, donde los subíndices tienen los siguientes significados: hg = graduado de escuela media superior, y cg = graduado universitario. En total, existen $n_1$ graduados de escuela media superior y $n_2$ graduados universitarios, para una muestra total de $n = n_1 + n_2$.

* Los datos de los resultados de la regresión provienen de N.M. Meltz, "Interstate and Interprovincial Differences in Union Density", *Industrial Relations*, vol. 28, núm. 2, 1989, pp. 142-158.

9.16. Para estudiar la tasa de crecimiento de la población de Belice de 1970 a 1992, Mukherjee *et al.* estimaron los siguientes modelos:*

$$\text{Modelo I:} \quad \widehat{\ln(\text{Pob})_t} = \underset{(781.25)}{4.73} + \underset{(54.71)}{0.024t}$$

$$\text{Modelo II:} \quad \widehat{\ln(\text{Pob})_t} = \underset{(2\,477.92)}{4.77} + \underset{(34.01)}{0.015t} - \underset{(-17.03)}{0.075D_t} + \underset{(25.54)}{0.011(D_t t)}$$

(los valores entre paréntesis son los estadísticos $t$)

donde Pob = población en millones; $t$ = variable de tendencia; $D_t = 1$ para observaciones que comenzaron en 1978 y 0 antes de 1978; y ln significa logaritmo natural.

*a*) En el modelo I, ¿cuál es la tasa de crecimiento de la población de Belice durante el periodo de muestra?

*b*) ¿Las tasas de crecimiento son estadísticamente distintas antes y después de 1978? ¿Cómo sabe? Si son diferentes, ¿cuáles son las tasas de crecimiento para 1972-1977 y 1978-1992?

## Ejercicios empíricos

9.17. Con los datos de la tabla 9.8 pruebe la hipótesis de que las varianzas de los errores en los dos subperiodos 1958-IV a 1966-III y 1966-IV a 1971-II son iguales.

9.18. Con la metodología analizada en el capítulo 8, compare la regresión no restringida (9.7.3) con la restringida (9.7.4); es decir, pruebe la validez de las restricciones impuestas.

9.19. En el ejemplo de la regresión (9.5.4) sobre ahorro-ingreso en Estados Unidos analizado en este capítulo, suponga que en lugar de utilizar los valores 0 y 1 para la variable dicótoma se emplea $Z_i = a + bD_i$, donde $D_i = 1$ y 0, $a = 2$ y $b = 3$. Compare sus resultados.

9.20. Siguiendo con la regresión (9.5.4) sobre ahorro-ingreso, suponga que asignó $D_i = 0$ para las observaciones hechas en el segundo periodo, y $D_i = 1$ para las observaciones del primer periodo. ¿Cómo cambiarían los resultados mostrados en (9.5.4)?

9.21. Con los datos de la tabla 9.2 considere el siguiente modelo:

$$\ln \text{Ahorro}_i = \beta_1 + \beta_2 \ln \text{Ingreso}_i + \beta_3 \ln D_i + u_i$$

donde ln significa logaritmo natural y $D_i = 1$ para 1970-1981 y 10 para 1982-1995.

*a*) ¿Cuál es el razonamiento en el que se basa la asignación de valores dicótomos, como se propone?

*b*) Estime el modelo anterior e interprete sus resultados.

*c*) ¿Cuáles son los valores del intercepto de la función ahorros en los dos subperiodos, y cómo los interpretaría?

9.22. Consulte las ventas trimestrales de aparatos electrodomésticos dadas en la tabla 9.3. Considere el siguiente modelo:

$$\text{Ventas}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + u_i$$

donde las variables dicótomas $D$ toman los valores de 1 y 0 para los trimestres II al IV.

*a*) Estime de manera individual el modelo anterior para las lavalozas, trituradores de basura y lavadoras.

*b*) ¿Cómo interpretaría los coeficientes estimados de las pendientes?

*c*) ¿Cómo utilizaría las $\alpha$ estimadas para desestacionalizar los datos sobre ventas de cada categoría de aparato?

---

* Chandan Mukherjee, Howard White y Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, Londres, 1998, pp. 372-375. Se adaptaron las notaciones al libro.

**TABLA 9.8**
**Elecciones presidenciales de Estados Unidos, 1916-2004**

| Obs. | Año | *V* | *W* | *D* | *G* | *I* | *N* | *P* |
|---|---|---|---|---|---|---|---|---|
| 1 | 1916 | 0.5168 | 0 | 1 | 2.229 | 1 | 3 | 4.252 |
| 2 | 1920 | 0.3612 | 1 | 0 | –11.46 | 1 | 5 | 16.535 |
| 3 | 1924 | 0.4176 | 0 | –1 | –3.872 | –1 | 10 | 5.161 |
| 4 | 1928 | 0.4118 | 0 | 0 | 4.623 | –1 | 7 | 0.183 |
| 5 | 1932 | 0.5916 | 0 | –1 | –14.9 | –1 | 4 | 7.069 |
| 6 | 1936 | 0.6246 | 0 | 1 | 11.921 | 1 | 9 | 2.362 |
| 7 | 1940 | 0.55 | 0 | 1 | 3.708 | 1 | 8 | 0.028 |
| 8 | 1944 | 0.5377 | 1 | 1 | 4.119 | 1 | 14 | 5.678 |
| 9 | 1948 | 0.5237 | 1 | 1 | 1.849 | 1 | 5 | 8.722 |
| 10 | 1952 | 0.446 | 0 | 0 | 0.627 | 1 | 6 | 2.288 |
| 11 | 1956 | 0.4224 | 0 | –1 | –1.527 | –1 | 5 | 1.936 |
| 12 | 1960 | 0.5009 | 0 | 0 | 0.114 | –1 | 5 | 1.932 |
| 13 | 1964 | 0.6134 | 0 | 1 | 5.054 | 1 | 10 | 1.247 |
| 14 | 1968 | 0.496 | 0 | 0 | 4.836 | 1 | 7 | 3.215 |
| 15 | 1972 | 0.3821 | 0 | –1 | 6.278 | –1 | 4 | 4.766 |
| 16 | 1976 | 0.5105 | 0 | 0 | 3.663 | –1 | 4 | 7.657 |
| 17 | 1980 | 0.447 | 0 | 1 | –3.789 | 1 | 5 | 8.093 |
| 18 | 1984 | 0.4083 | 0 | –1 | 5.387 | –1 | 7 | 5.403 |
| 19 | 1988 | 0.461 | 0 | 0 | 2.068 | –1 | 6 | 3.272 |
| 20 | 1992 | 0.5345 | 0 | –1 | 2.293 | –1 | 1 | 3.692 |
| 21 | 1996 | 0.5474 | 0 | 1 | 2.918 | 1 | 3 | 2.268 |
| 22 | 2000 | 0.50265 | 0 | 0 | 1.219 | 1 | 8 | 1.605 |
| 23 | 2004 | 0.51233 | 0 | 1 | 2.69 | –1 | 1 | 2.325 |

*Notas:*
Año Año electoral
*V* Participación titular del voto bipartidista para la presidencia.
*W* Variable indicadora (1 para las elecciones de 1920, 1944 y 1948; 0 en cualquier otro caso).
*D* Variable indicadora (1 si un candidato demócrata compite en las elecciones, −1 si un candidato republicano compite en las elecciones; 0 en cualquier otro caso).
*G* Tasa de crecimiento del PIB *per cápita* real en los primeros tres trimestres del año electoral.
*I* Variable indicadora (1 si hay un titular demócrata al momento de las elecciones y –1 si hay un titular republicano).
*N* Número de trimestres en los primeros 15 trimestres de la administración en los que la tasa de crecimiento del PIB *per cápita* real es mayor que 3.2%.
*P* Valor absoluto de la tasa de crecimiento del deflactor del PIB en los primeros 15 trimestres de la administración.

9.23. Vuelva a estimar el modelo del ejercicio 9.22 añadiendo la siguiente regresora: gastos en bienes duraderos.

*a*) ¿Existe alguna diferencia entre los resultados de la regresión obtenidos en el ejercicio 9.22 y los de este ejercicio? De ser así, ¿qué explica la diferencia?

*b*) Si existiera estacionalidad en los datos del gasto en bienes duraderos, ¿cómo la explicaría?

9.24. La tabla 9.9 proporciona datos sobre las elecciones presidenciales de Estados Unidos de 1916 a 2004.*

*a*) Con los datos de la tabla 9.9 elabore un modelo adecuado para predecir la proporción correspondiente al Partido Demócrata del voto bipartidista para la presidencia.

*b*) ¿Cómo utilizaría este modelo para predecir el resultado de una elección presidencial?

* Ray Fair, de la Universidad de Yale, recopiló originalmente estos datos; Fair ha pronosticado los resultados de las elecciones presidenciales durante varios años. Los datos se reproducen de la obra de Samprit Chatterjee, Ali S. Hadi y Petram Price, *Regression Analysis by Example*, 3a. ed., John Wiley & Sons, Nueva York, 2000, pp. 150-151, y se actualizaron de http://fairmodel.econ.yale.edu/rayfair/pdf/2006CHTM.HTM

*c*) Chatterjee *et al.* propusieron considerar el siguiente modelo tentativo para predecir las elecciones presidenciales:

$$V = \beta_0 + \beta_1 I + \beta_2 D + \beta_3 W + \beta_4 (GI) + \beta_5 P + \beta_6 N + u$$

Estime este modelo y comente los resultados respecto de los resultados del modelo que había propuesto.

**9.25.** Consulte la regresión (9.6.4). Pruebe la hipótesis de que la tasa de crecimiento de los ingresos promedio por hora respecto de la escolaridad difiere según el sexo y la raza. (*Sugerencia:* Utilice variables dicótomas multiplicativas.)

**9.26.** Consulte la regresión (9.3.1). ¿Cómo modificaría el modelo para saber si existe alguna interacción entre las variables sexo y región de residencia? Presente los resultados basados en este modelo y compárelos con los dados en (9.3.1).

**9.27.** En el modelo $Y_i = \beta_1 + \beta_2 D_i + u_i$, sea $D_i = 0$ para las primeras 40 observaciones y $D_i = 1$ para las restantes 60 observaciones. Sabe que $u_i$ tiene media cero y varianza de 100. ¿Cuáles son los valores medios y las varianzas para los dos conjuntos de observaciones?*

**9.28.** Consulte la regresión de ahorro-ingreso analizada en este capítulo. Como modelo alterno a (9.5.1), considere el siguiente:

$$\ln Y_t = \beta_1 + \beta_2 D_t + \beta_3 X_t + \beta_4 (D_t X_t) + u_t$$

donde $Y$ representa los ahorros y $X$ el ingreso.

*a*) Estime el modelo anterior y compare los resultados con los de (9.5.4). ¿Cuál es mejor modelo?

*b*) ¿Cómo interpretaría el coeficiente de la variable dicótoma en este modelo?

*c*) Como veremos en el capítulo sobre la heteroscedasticidad, a menudo una transformación logarítmica de la variable dependiente reduce la heteroscedasticidad en los datos. Vea si éste es el caso del ejemplo presente con la regresión del logaritmo de $Y$ sobre $X$ para los dos periodos, y observe si las varianzas de error estimadas para los dos periodos son estadísticamente iguales. Si lo son, se puede utilizar la prueba de Chow para agrupar los datos de la manera indicada en el capítulo.

9.29. Consulte el ejemplo de los trabajadores asalariados de India (sección 9.12) y los datos de la tabla 9.7.† Como recordatorio, las variables se definen como sigue:

IS = ingreso por salario semanal en rupias
Edad = edad en años
$D_{\text{sexo}}$ = 1 para trabajadores y 0 para trabajadoras
$DE_2$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad hasta primaria
$DE_3$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad hasta secundaria
$DE_4$ = variable dicótoma que toma el valor de 1 para trabajadores con nivel de escolaridad superior al nivel secundaria
$DPT$ = variable dicótoma que toma el valor de 1 para trabajadores con empleo permanente y 0 para trabajadores eventuales

La categoría de referencia es trabajadores sin educación primaria y con empleo eventual.

* Este ejemplo se adaptó de Peter Kennedy, *A Guide to Econometrics*, 4a. ed., MIT Press, Cambridge, Massachusetts, 1998, p. 347.

† Los datos provienen de *Econometrics and Data Analysis for Developing Countries*, de Chandan Mukherjee, Howard White y Marc Wuyts, Routledge Press, Londres, 1998, en el apéndice.

En la sección 9.12 se crearon términos de interacción entre las variables de escolaridad ($DE_2$, $DE_3$ y $DE_4$) y la variable de sexo ($D_{\text{sexo}}$). ¿Qué sucede si se crean términos de interacción entre las variables dicótomas de escolaridad y la variable dicótoma de trabajadores permanentes ($DPT$)?

*a*) Estime el modelo para predecir ln IS que contenga variables dicótomas edad, sexo y escolaridad, y tres nuevos términos de interacción: $DE_2 \times DPT$, $DE_3 \times DPT$ y $DE_4 \times DPT$. ¿Parece que existe algún efecto significativo de interacción entre los nuevos términos?

*b*) ¿Hay alguna diferencia significativa entre los trabajadores con nivel de escolaridad hasta primaria y los trabajadores que no tienen educación primaria? Evalúe esto respecto tanto de la variable dicótoma de escolaridad como del término de interacción y explique los resultados. ¿Qué sucede con la diferencia entre los trabajadores que tienen educación secundaria y los que no tienen el nivel de educación primaria? ¿Y con la diferencia entre quienes tienen nivel de escolaridad superior a secundaria, en comparación con los que no tienen educación primaria?

*c*) Ahora evalúe los resultados de eliminar las variables dicótomas del modelo. ¿Cambia la significancia de los términos de interacción?

# Apéndice 9A

## Regresión semilogarítmica con regresora dicótoma

En la sección 9.10 observamos que en los modelos del tipo

$$\ln Y_i = \beta_1 + \beta_2 D_i \tag{1}$$

el cambio relativo en $Y$ (es decir, la semielasticidad), respecto de la regresora dicótoma que toma los valores de 1 o 0, se obtiene como (el antilogaritmo de la $\beta_2$ estimada) − 1 por 100; es decir, como

$$(e^{\hat{\beta}_2} - 1) \times 100 \tag{2}$$

La prueba es como sigue: como la función log y exp ($= e$) son inversas, (1) se expresa como:

$$\ln Y_i = \beta_1 + \ln(e^{\beta_2 D_i}) \tag{3}$$

Ahora bien, cuando $D = 0$, $e^{\beta_2 D_i} = 1$, y cuando $D = 1$, $e^{\beta_2 D_i} = e^{\beta_2}$. Por consiguiente, al pasar del estado 0 al 1, ln $Y_i$ cambia por $(e^{\beta_2} - 1)$. Pero una variación en el logaritmo de una variable es un cambio relativo, el cual, después de la multiplicación por 100, se convierte en un cambio porcentual. Por tanto, ese cambio porcentual es $(e^{\beta_2} - 1) \times 100$, como se afirmaba. (*Nota:* $\ln_e e = 1$, es decir, el log de $e$, base $e$, es 1; al igual que el logaritmo de 10, base 10, es 1. Recuerde que el logaritmo base $e$ se llama logaritmo *natural,* y que el logaritmo base 10 se llama logaritmo *común*.)

# Flexibilización de los supuestos del modelo clásico

# Parte 2

En la parte 1 consideramos extensamente el modelo clásico de regresión lineal normal y vimos cómo utilizarlo para manejar dos problemas de inferencia estadística, a saber, la estimación y las pruebas de hipótesis, así como el problema de predicción. Pero recuerde que este modelo está basado en diversos supuestos simplificadores, que son los siguientes:

Supuesto 1. El modelo de regresión es lineal en los parámetros.

Supuesto 2. Los valores de las regresoras, las *X*, son fijos, o los valores de *X* son independientes del término de error. Aquí, esto significa que se requiere covarianza cero entre $u_i$ y cada variable *X*.

Supuesto 3. Para *X* dadas, el valor medio de la perturbación $u_i$ es cero.

Supuesto 4. Para *X* dadas, la varianza de $u_i$ es constante u homoscedástica.

Supuesto 5. Para *X* dadas, no hay autocorrelación, o correlación serial, entre las perturbaciones.

Supuesto 6. El número de observaciones *n* debe ser mayor que el número de parámetros por estimar.

Supuesto 7. Debe haber variación suficiente entre los valores de las variables *X*.

También se incluyen los siguientes tres supuestos en esta parte del texto:

Supuesto 8. No hay colinealidad exacta entre las variables *X*.

Supuesto 9. El modelo está correctamente especificado, por lo que no hay sesgo de especificación.

Supuesto 10. El término estocástico (de perturbación) $u_i$ está normalmente distribuido.

Antes de proseguir, conviene señalar que la mayoría de los libros de texto enumeran menos de 10 supuestos. Por ejemplo, los supuestos 6 y 7 se dan por obvios en vez de declararlos explícitamente. Decidimos hacerlos explícitos porque parece razonable distinguir entre los supuestos requeridos para que los MCO tengan las propiedades estadísticas deseables (como MELI) y las condiciones requeridas para que los MCO sean útiles. Por ejemplo, los estimadores de MCO son estimadores MELI aunque no se satisfaga el supuesto 7. Pero en ese caso, los errores estándar de los estimadores de MCO serán grandes en comparación con sus coeficientes (es decir, las razones

$t$ serán pequeñas), con lo cual se dificulta evaluar la contribución de una o más regresoras a la suma de cuadrados explicada.

Como menciona Wetherill, en la práctica surgen dos tipos de problemas importantes al aplicar el modelo clásico de regresión lineal: 1) los debidos a supuestos sobre la especificación del modelo y sobre las perturbaciones $u_i$, y 2) los debidos a los supuestos sobre los datos.[1] En la primera categoría están los supuestos 1, 2, 3, 4, 5, 9 y 10. Los de la segunda categoría son los supuestos 6, 7 y 8. Además, los problemas en los datos, como observaciones atípicas (o inusuales) y errores de medición, también se encuentran en la segunda categoría.

Respecto de los problemas por los supuestos sobre las perturbaciones y especificaciones del modelo, surgen tres grandes interrogantes: 1) ¿Qué gravedad reviste alejarse de un supuesto particular para que de verdad importe? Por ejemplo, si las $u_i$ no siguen una distribución exactamente normal, ¿qué nivel de alejamiento de este supuesto puede aceptarse sin que se destruya la propiedad MELI de los estimadores de MCO? 2) ¿Cómo averiguar si en verdad se violó un supuesto particular en un caso concreto? Por consiguiente, ¿cómo verificamos que las perturbaciones estén normalmente distribuidas en una aplicación dada? Ya estudiamos las pruebas de normalidad del **estadístico $A^2$ de Anderson-Darling** y de **Jarque-Bera**. 3) ¿Qué medidas correctivas podemos emprender si uno o más supuestos son falsos? Por ejemplo, si encontramos que el supuesto de homoscedasticidad en una aplicación es falso, ¿qué hacer?

Respecto de los problemas atribuibles a supuestos sobre los datos, también se presentan interrogantes similares. 1) ¿Qué tan grave es un problema particular? Por ejemplo, ¿es la multicolinealidad tan severa que dificulta mucho la estimación y la inferencia? 2) ¿Cómo se mide la severidad del problema de datos? Por ejemplo, ¿cómo decidimos si la inclusión o la exclusión de una observación u observaciones que pueden representar observaciones atípicas constituirán una diferencia sustancial en el análisis? 3) ¿Es posible remediar con facilidad algunos problemas de información? Por ejemplo, ¿se puede acceder a los datos originales para encontrar las fuentes de los errores de medición en los datos?

Por desgracia, no hay respuestas satisfactorias a todas estas preguntas. Lo que haremos en la parte 2 es mirar algunos de los supuestos en forma más crítica, aunque no todos serán objeto de un escrutinio completo. En particular, no analizaremos a fondo los supuestos 2, 3 y 10, por las siguientes razones:

### *Supuesto 2: Regresoras fijas y regresoras estocásticas*

Recuerde que el análisis de regresión se basa en el supuesto de que las regresoras son no estocásticas y suponen valores fijos en muestreo repetido. Existe una buena razón para esta estrategia. A diferencia de los científicos de las ciencias físicas, como mencionamos en el capítulo 1, los economistas generalmente no tienen control sobre sus datos, dependen con mayor frecuencia de datos secundarios, es decir, información recopilada por otros agentes, como el gobierno y organizaciones privadas. Por consiguiente, la estrategia práctica es suponer que, para el problema en curso, los valores de las variables explicativas están dados aunque las variables mismas sean intrínsecamente estocásticas o aleatorias. Por tanto, los resultados del análisis de regresión están condicionados a estos valores dados.

Pero suponga que no es posible considerar las $X$ realmente no estocásticas o fijas. Es el caso de las **regresoras aleatorias** o **estocásticas**. Ahora la situación se complica. Las $u_i$, por supuesto, son estocásticas. Si las $X$ también lo son, se debe especificar la distribución tanto de las $X$ como

[1] G. Barrie Wetherill, *Regression Analysis with Applications*, Chapman y Hall, Nueva York, 1986, pp. 14-15.

de las $u_i$. Si estamos dispuesto a considerar el supuesto 2 (es decir, que las $X$, aunque aleatorias, están distribuidas independientemente, o por lo menos no están correlacionadas, con las $u_i$), para todo fin práctico se puede trabajar como si las $X$ fueran no estocásticas. Como señala Kmenta:

> Así, *la no consideración o la exclusión del supuesto de que X es no estocástica y su reemplazo por el supuesto de que X es estocástica aunque independiente de [u] no cambia las propiedades deseables y la factibilidad de la estimación de mínimos cuadrados.*[2]

Por consiguiente, conservaremos el supuesto 2 hasta que abordemos el tema de los modelos de ecuaciones simultáneas, en la parte 4.[3] Además, presentaremos un análisis breve de las regresoras no estocásticas en el capítulo 13.

### *Supuesto 3: Valor de la media de $u_i$ igual a cero*

Recuerde el modelo de regresión lineal con $k$ variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

Supongamos ahora que

$$E(u_i | X_{2i}, X_{3i}, \ldots, X_{ki}) = w \tag{2}$$

donde $w$ es una constante; observe que en el modelo estándar $w = 0$, pero ahora puede ser cualquier constante.

Tomamos la esperanza condicional de (1) y obtenemos

$$\begin{aligned} E(Y_i | X_{2i}, X_{3i}, \ldots, X_{ki}) &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + w \\ &= (\beta_1 + w) + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \\ &= \alpha + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \end{aligned} \tag{3}$$

donde $\alpha = (\beta_1 + w)$ y donde, al tomar las esperanzas, se debe recordar que las $X$ se tratan como constantes. (¿Por qué?)

Por consiguiente, si no se satisface el supuesto 3, no podemos estimar el intercepto original $\beta_1$; lo que se obtiene es $\alpha$, el cual contiene a $\beta_1$ y a $E(u_i) = w$. En resumen, obtenemos una estimación *sesgada* de $\beta_1$.

Pero, como mencionamos en diversas ocasiones, en muchas situaciones prácticas el intercepto, $\beta_1$, es de poca importancia; los parámetros con mayor significado son los coeficientes de pendiente, que permanecen inalterados aunque se viole el supuesto 3.[4] Además, en muchas aplicaciones el término del intercepto no tiene interpretación alguna.

---

[2] Jan Kmenta, *Elements of Econometrics*, 2a. ed., Macmillan, Nueva York, 1986, p. 338. (Cursivas en el original.)

[3] Aquí debe observarse un punto técnico. En lugar del supuesto sólido de que las $X$ y $u$ son independientes, puede servir el supuesto más débil de que los valores de las variables $X$ y las $u$ no están correlacionados contemporáneamente (es decir, en el mismo momento). En este caso, los estimadores de MCO pueden estar sesgados pero son **consistentes**, es decir, a medida que el tamaño de la muestra se incrementa indefinidamente, los estimadores convergen hacia sus valores verdaderos. Sin embargo, si las $X$ y $u$ están correlacionadas contemporáneamente, los estimadores de MCO son sesgados al igual que inconsistentes. En el capítulo 17 veremos que, en esta situación, algunas veces el método de **variables instrumentales** sirve para obtener estimadores consistentes.

[4] Es muy importante mencionar que esta afirmación es válida sólo si $E(u_i) = w$ para cada $i$. Sin embargo, si $E(u_i) = w_i$, es decir, una constante diferente para cada $i$, los coeficientes de pendiente parciales pueden ser sesgados lo mismo que inconsistentes. En este caso, la violación del supuesto 3 será relevante. Para una demostración y mayores detalles, véase Peter Schmidt, *Econometrics*, Marcel Dekker, Nueva York, 1976, pp. 36-39.

### *Supuesto 10: Normalidad de u*

Este supuesto no es esencial si el objetivo es solamente la estimación. Como vimos en el capítulo 3, los estimadores de MCO son MELI sin importar que las $u_i$ estén normalmente distribuidas o no. Con el supuesto de normalidad, sin embargo, es posible establecer que los estimadores de MCO de los coeficientes de regresión siguen la distribución normal, que $(n-k)\hat{\sigma}/\sigma^2$ sigue la distribución $\chi^2$ y que pueden utilizarse las pruebas $t$ y $F$ para verificar diversas hipótesis estadísticas, sin importar el tamaño de la muestra.

Pero, ¿qué sucede si las $u_i$ no están normalmente distribuidas? Entonces se toma en cuenta la siguiente extensión del teorema central del límite; recuerde que fue el teorema central del límite, en primera instancia, el que empleamos para justificar el supuesto de normalidad:

> Si las perturbaciones [$u_i$] son independientes y están idénticamente distribuidas con media cero y varianza [constante] $\sigma^2$, y si las variables explicativas son constantes en muestras repetidas, los estimadores de MC[O] de los coeficientes son asintóticamente normales con medias iguales a las $\beta$ correspondientes.[5]

Por consiguiente, los procedimientos de prueba usuales —las pruebas $t$ y $F$— son aún válidas *asintóticamente*, es decir, en muestras grandes, pero no en muestras finitas o pequeñas.

El hecho de que los estimadores de MCO sigan una distribución normal asintótica (según el supuesto de varianza homoscedástica y valores fijos de $X$) aunque las perturbaciones no tengan distribución normal es de poca ayuda para los analistas económicos, que pocas veces disponen de datos de muestras grandes. Por tanto, el supuesto de normalidad adquiere gran importancia para los fines de pruebas de hipótesis y predicción. Entonces, teniendo en mente los problemas de estimación y de pruebas de hipótesis, y debido a que las muestras pequeñas son la regla más que la excepción en la mayoría de los análisis económicos, debemos mantener el supuesto de normalidad.[6] (Pero véase el capítulo 13, sección 13.12.)

Por supuesto, esto significa que, cuando se trata de una muestra finita, se debe realizar la prueba explícita del supuesto de normalidad. Ya consideramos las **pruebas de Anderson-Darling** y **Jarque-Bera** de normalidad. Se sugiere al lector aplicar éstas u otras pruebas de normalidad a los residuos de la regresión. Debe tener en mente que, en muestras finitas sin el supuesto de normalidad, los estadísticos usuales $t$ y $F$ pueden no seguir las distribuciones $t$ y $F$.

Quedan los supuestos 1, 4, 5, 6, 7, 8 y 9. Los supuestos 6, 7 y 8 están estrechamente interrelacionados y se analizan en el capítulo sobre multicolinealidad (10). El supuesto 4 se estudia en el capítulo sobre heteroscedasticidad (11); el supuesto 5, en el capítulo sobre autocorrelación (12) y el supuesto 9, en el capítulo sobre especificación de modelos y prueba de diagnóstico (13). Debido a lo especializado de la naturaleza y a las exigencias matemáticas, analizaremos el supuesto 1 como tema especial en la parte 3 (capítulo 14).

Por razones pedagógicas, en cada uno de esos capítulos seguiremos un esquema común, a saber, 1) identificar la naturaleza del problema, 2) examinar sus consecuencias, 3) sugerir métodos para detectarlo y 4) considerar medidas correctivas de forma que produzcan estimadores con las propiedades estadísticas deseables analizadas en la parte 1.

---

[5] Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, p. 240. Debe mencionarse que los supuestos de $X$ fijas y $\sigma^2$ constantes son cruciales para este resultado.

[6] A propósito, observe que los efectos del incumplimiento de la normalidad y temas relacionados suele analizarse en el tema de **estimación robusta** en la teoría, lo cual trasciende el alcance de este libro.

Cabe anotar: como ya mencionamos, no hay respuestas satisfactorias a todos los problemas que surgen de la violación de los supuestos del MCRL. Además, puede haber más de una solución a un problema particular y a menudo no es claro cuál es la mejor. Además, en una aplicación particular puede haber más de una violación al MCRL. Así, el sesgo de especificación, la multicolinealidad y la heteroscedasticidad pueden coexistir en una aplicación, y no existe una prueba única omnipotente que resuelva todos los problemas al mismo tiempo.[7] Más aún, una prueba específica, que pudo ser muy común en alguna época, quizá ya no se aplique porque se descubrió que tenía alguna falla. Pero así avanza la ciencia. La econometría no es la excepción.

[7] Esto no es por falta de intentos. Véase A.K. Bera y C.M. Jarque, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence", *Economic Letters*, vol. 7, 1981, pp. 313-318.

# Capítulo 10

# Multicolinealidad: ¿qué pasa si las regresoras están correlacionadas?

> No hay una expresión más errónea, tanto en los libros de texto de econometría como en la bibliografía aplicada, que la de "problema de multicolinealidad". Es un hecho que muchas variables explicativas presentan un alto grado de colinealidad; asimismo, resulta muy claro que existen diseños experimentales $\mathbf{X}'\mathbf{X}$ (es decir, matriz de datos) que serían mucho más convenientes que los diseños que proporciona la experimentación natural (es decir, la muestra disponible). No obstante, no es nada constructivo quejarse de la aparente malevolencia de la naturaleza, y los remedios *ad hoc* para un mal diseño —como una regresión por pasos o una regresión en cadena— pueden ser desastrosamente inapropiados. Es mejor aceptar de plano que los datos que no se recopilaron mediante experimentos diseñados a veces no proporcionan mucha información sobre los parámetros de interés.[1]

El supuesto 8 del *modelo clásico de regresión lineal* (MCRL) plantea que no existe **multicolinealidad** entre las regresoras incluidas en el modelo de regresión. En este capítulo consideramos en forma crítica el supuesto de no multicolinealidad en busca de respuestas a las siguientes preguntas:

1. ¿Cuál es la naturaleza de la multicolinealidad?
2. ¿Es la multicolinealidad realmente un problema?
3. ¿Cuáles son sus consecuencias prácticas?
4. ¿Cómo se detecta?
5. ¿Qué medidas pueden tomarse para aliviar el problema de multicolinealidad?

En este capítulo también analizaremos el supuesto 6 del MCRL, a saber, que el número de observaciones en la muestra debe ser mayor que el de regresoras, así como el supuesto 7, que requiere una variabilidad suficiente en los valores de las regresoras, en vista de que ambos

[1] Edward E. Leamer, "Model Choice and Specification Analysis", en Zvi Griliches y Michael D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1, North Holland Publishing Company, Ámsterdam, 1983, pp. 300-301.

están estrechamente relacionados con el supuesto de la multicolinealidad. Arthur Goldberger denominó al supuesto 6 el problema de la **micronumerosidad**,[2] lo cual simplemente significa un tamaño pequeño de muestra.

## 10.1 Naturaleza de la multicolinealidad

El término *multicolinealidad* se atribuye a Ragnar Frisch.[3] Originalmente, designaba una relación lineal "perfecta" o exacta entre algunas o todas las variables explicativas de un modelo de regresión.[4] Para la regresión con $k$ variables que incluye las variables explicativas $X_1, X_2, \ldots, X_k$ (donde $X_1 = 1$ para todas las observaciones de forma que den cabida al término del intercepto), se dice que existe una relación lineal exacta si se satisface la siguiente condición:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0 \qquad \textbf{(10.1.1)}$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$, son constantes tales que no todas son simultáneamente iguales a cero.[5]

Hoy en día, sin embargo, el término multicolinealidad incluye el caso de multicolinealidad perfecta, como lo indica (10.1.1) y también el caso en el cual hay $X$ variables intercorrelacionadas pero no en forma perfecta, de la siguiente manera:[6]

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_2 X_k + v_i = 0 \qquad \textbf{(10.1.2)}$$

donde $v_i$ es un término de error estocástico.

Para apreciar la diferencia entre multicolinealidad *perfecta* y multicolinealidad *menos que perfecta* suponga, por ejemplo, que $\lambda_2 \neq 0$. Entonces, (10.1.1) se escribe como

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \cdots - \frac{\lambda_k}{\lambda_2} X_{ki} \qquad \textbf{(10.1.3)}$$

que muestra la forma como $X_2$ está exactamente relacionada de manera lineal con otras variables, o cómo se deriva de una combinación lineal de otras variables $X$. En esta situación, el coeficiente de correlación entre la variable $X_2$ y la combinación lineal del lado derecho de (10.1.3) está obligado a ser igual a uno.

En forma similar, si $\lambda_2 \neq 0$, la ecuación (10.1.2) se escribe como

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \cdots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{1}{\lambda_2} v_i \qquad \textbf{(10.1.4)}$$

lo cual muestra que $X_2$ no es una combinación lineal exacta de otras $X$ porque está determinada también por el término de error estocástico $v_i$.

---

[2] Véase su obra *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991, p. 249.

[3] Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Institute of Economics, Oslo University, publ. núm. 5, 1934.

[4] En estricto sentido, la *multicolinealidad* se refiere a la existencia de más de una relación lineal exacta, y *colinealidad*, a la existencia de una sola relación lineal. Pero esta distinción pocas veces se mantiene en la práctica, y se hace entonces referencia a multicolinealidad en ambos casos.

[5] Las posibilidades de obtener una muestra de valores en la cual las regresoras estén relacionadas en esta forma son sin duda muy pocas en la práctica, excepto por diseño, cuando, por ejemplo, el número de observaciones es menor que el número de regresoras, o si se cae en la "trampa de la variable dicótoma", como analizamos en e1 capítulo 9. Véase el ejercicio 10.2.

[6] Si sólo hay dos variables explicativas, la *intercorrelación* se mide con el coeficiente de orden cero o de correlación simple. Pero si hay más de dos variables $X$, la intercorrelación se mide por los coeficientes de correlación parcial o por el coeficiente de correlación múltiple $R$ de una variable $X$, y se consideran en conjunto todas las demás variables $X$.

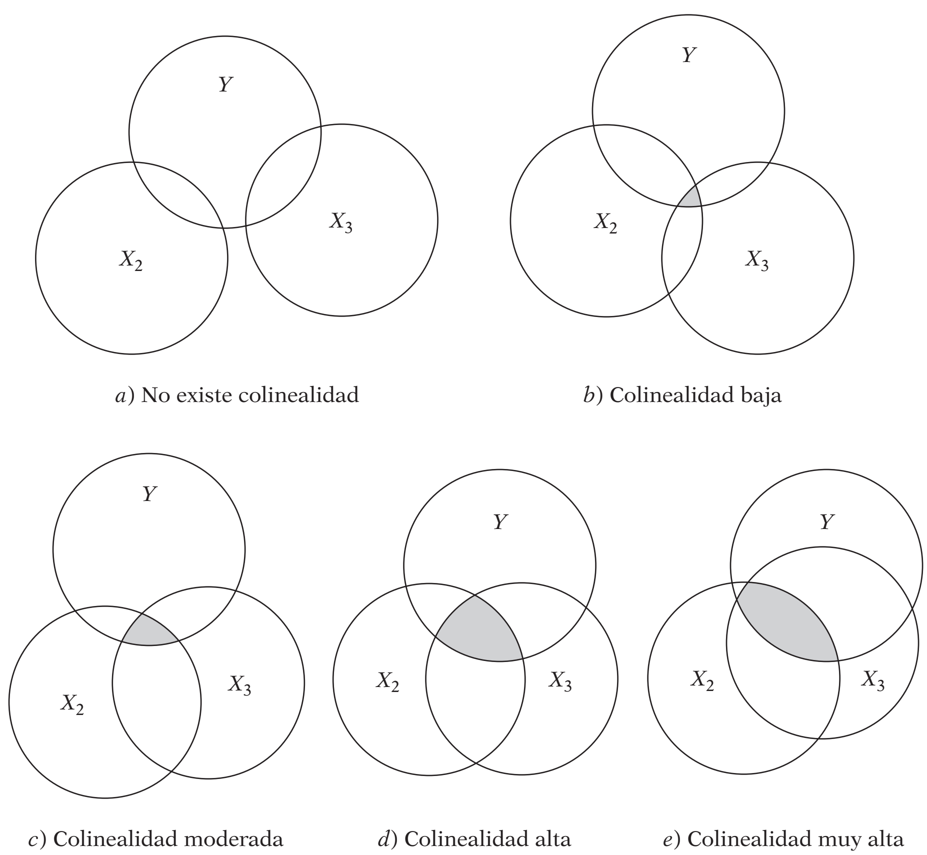
Como ejemplo numérico, considere la siguiente información hipotética:

| $X_2$ | $X_3$ | $X_3^*$ |
|---|---|---|
| 10 | 50 | 52 |
| 15 | 75 | 75 |
| 18 | 90 | 97 |
| 24 | 120 | 129 |
| 30 | 150 | 152 |

Es evidente que $X_{3i} = 5X_{2i}$. Por consiguiente, hay colinealidad perfecta entre $X_2$ y $X_3$, pues el coeficiente de correlación $r_{23}$ es la unidad. La variable $X_3^*$ se creó de $X_3$ agregándole simplemente los siguientes números, tomados de una tabla de números aleatorios: 2, 0, 7, 9, 2. Ahora ya no hay multicolinealidad perfecta entre $X_2$ y $X_3^*$. Sin embargo, las dos variables están muy correlacionadas, pues los cálculos indicarán que el coeficiente de correlación entre ellas es 0.9959.

El método algebraico anterior para el problema de la multicolinealidad se expresa concisamente mediante un diagrama de Ballentine (recuerde la figura 3.8 que se reproduce en la figura 10.1). En esta figura los círculos $Y$, $X_2$ y $X_3$ representan las variaciones en $Y$ (la variable dependiente) y en $X_2$ y $X_3$ (las variables explicativas). El grado de colinealidad se mide por la magnitud de la intersección (área sombreada) de los círculos $X_2$ y $X_3$. En la figura 10.1*a*) no hay intersección entre $X_2$ y $X_3$, y, por tanto, no hay colinealidad. En las figuras 10.1*b*) a 10.1*e*), el grado de colinealidad va de "bajo" a "alto": entre mayor sea la intersección entre $X_2$ y $X_3$ (es decir, entre

**FIGURA 10.1**
Gráfico de Ballentine de multicolinealidad.



mayor sea el área sombreada), mayor será el grado de colinealidad. En el extremo, si $X_2$ y $X_3$ estuvieran superpuestos completamente (o si $X_2$ estuviera por completo dentro de $X_3$, o viceversa), la colinealidad sería perfecta.

A propósito, observe que la multicolinealidad, como la definimos, se refiere sólo a relaciones lineales entre las variables $X$. Este concepto no aplica a las relaciones no lineales entre ellas. Por ejemplo, considere el siguiente modelo de regresión:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad \textbf{(10.1.5)}$$

donde, digamos, $Y =$ costo total de producción y $X =$ producción. Las variables $X_i^2$ (producción al cuadrado) y $X_i^3$ (producción al cubo) por supuesto están funcionalmente relacionadas con $X_i$, pero la relación es no lineal. De manera estricta, por consiguiente, modelos como (10.1.5) no violan el supuesto de no multicolinealidad. Sin embargo, en aplicaciones concretas, el coeficiente de correlación medido de forma convencional demostrará que $X_i$, $X_i^2$ y $X_i^3$ están altamente correlacionadas, lo cual, como mostraremos, dificultará estimar los parámetros de (10.1.5) con mayor precisión (es decir, con errores estándar pequeños).

¿Por qué supone el modelo clásico de regresión lineal que no hay multicolinealidad entre las $X$? El razonamiento es el siguiente: **Si la multicolinealidad es perfecta en el sentido de (10.1.1), los coeficientes de regresión de las variables $X$ son indeterminados, y sus errores estándar, infinitos. Si la multicolinealidad es menos que perfecta, como sucede en (10.1.2), los coeficientes de regresión, aunque sean determinados, poseen grandes errores estándar (en relación con los coeficientes mismos), lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud.** Las pruebas de estas afirmaciones se presentan en las siguientes secciones.

Existen diversas fuentes de multicolinealidad. Como afirman Montgomery y Peck, la multicolinealidad puede deberse a los siguientes factores:[7]

1. *El método de recolección de información.* Por ejemplo, la obtención de muestras en un intervalo limitado de valores tomados por las regresoras en la población.

2. *Restricciones en el modelo o en la población objeto de muestreo*. Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso ($X_2$) y el tamaño de las viviendas ($X_3$) hay una restricción física en la población, pues las familias con ingresos más altos suelen habitar viviendas más grandes que las familias con ingresos más bajos.

3. *Especificación del modelo.* Por ejemplo, la adición de términos polinomiales a un modelo de regresión, en especial cuando el rango de la variable $X$ es pequeño.

4. *Un modelo sobredeterminado*. Esto sucede cuando el modelo tiene más variables explicativas que el número de observaciones. Esto puede suceder en investigación médica, donde en ocasiones hay un número reducido de pacientes sobre quienes se reúne información respecto de un gran número de variables.

Otra razón para la multicolinealidad, sobre todo en los datos de series de tiempo, puede ser que las regresoras del modelo compartan una *tendencia común*; es decir, que todas aumenten o disminuyan a lo largo del tiempo. Por tanto, en la regresión del gasto de consumo sobre el ingreso, la riqueza y la población, las regresoras ingreso, riqueza y población tal vez todas crezcan con el tiempo a una tasa aproximadamente igual, con lo cual se presentaría la colinealidad entre dichas variables.

---

[7] Douglas Montgomery y Elizabeth Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, Nueva York, 1982, pp. 289-290. Véase también R.L. Mason, R.F. Gunst y J.T. Webster, "Regression Analysis and Problems of Multicollinearity," *Communications in Statistics A*, vol. 4, núm. 3, 1975, pp. 277-292; R. F. Gunst y R.L. Mason, "Advantages of Examining Multicollinearities in Regression Analysis", *Biometrics*, vol. 33, 1977, pp. 249-260.

## 10.2 Estimación en presencia de multicolinealidad perfecta

Ya establecimos que, en el caso de multicolinealidad perfecta, los coeficientes de regresión permanecen indeterminados y sus errores estándar son infinitos. Esto se demuestra fácilmente en términos del modelo de regresión con tres variables. Con la forma de desviación, en la cual todas las variables se expresan como desviaciones de sus medias muestrales, se escribe el modelo de regresión con tres variables como

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{10.2.1}$$

Ahora, del capítulo 7, obtenemos

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.7}$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.8}$$

Suponga que $X_{3i} = \lambda X_{2i}$, donde $\lambda$ es una constante diferente de cero (por ejemplo, 2, 4, 1.8, etc.). Si sustituimos esto en (7.4.7) obtenemos

$$\begin{aligned}\hat{\beta}_2 &= \frac{\left(\sum y_i x_{2i}\right)\left(\lambda^2 \sum x_{2i}^2\right) - \left(\lambda \sum y_i x_{2i}\right)\left(\lambda \sum x_{2i}^2\right)}{\left(\sum x_{2i}^2\right)\left(\lambda^2 \sum x_{2i}^2\right) - \lambda^2 \left(\sum x_{2i}^2\right)^2} \\ &= \frac{0}{0}\end{aligned} \tag{10.2.2}$$

que es una expresión indeterminada. El lector puede verificar que $\hat{\beta}_3$ también es indeterminada.[8]

¿Por qué obtenemos el resultado que aparece en (10.2.2)? Recuerde el significado de $\hat{\beta}_2$: da la tasa de cambio en el valor promedio de $Y$ a medida que $X_2$ cambia en una unidad, manteniendo $X_3$ constante. Pero si $X_3$ y $X_2$ son perfectamente colineales, no hay forma de que $X_3$ se mantenga constante: a medida que $X_2$ cambia, también lo hace $X_3$ por el factor $\lambda$. Esto significa, entonces, que no hay forma de desenredar las influencias separadas de $X_2$ y $X_3$ de la muestra dada: para fines prácticos, $X_2$ y $X_3$ son indistinguibles. En la econometría aplicada, este problema ocasiona mucho daño, pues la idea consiste en separar los efectos parciales de cada $X$ sobre la variable dependiente.

Para ver esto de otra forma, sustituya $X_{3i} = \lambda X_{2i}$ en (10.2.1) y obtendrá lo siguiente [véase también (7.1.12)]:

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3(\lambda x_{2i}) + \hat{u}_i \\ &= (\hat{\beta}_2 + \lambda\hat{\beta}_3) x_{2i} + \hat{u}_i \\ &= \hat{\alpha} x_{2i} + \hat{u}_i \end{aligned} \tag{10.2.3}$$

donde

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) \tag{10.2.4}$$

[8] Otra forma de ver esto es la siguiente: por definición, el coeficiente de correlación entre $X_2$ y $X_3$, $r_{23}$, es $\sum x_{2i} x_{3i} / \sqrt{\sum x_{2i}^2 \sum x_{3i}^2}$. Si $r_{23}^2 = 1$, es decir, si existe colinealidad perfecta entre $X_2$ y $X_3$, el denominador de (7.4.7) será cero, lo que imposibilita la estimación de $\beta_2$ (o de $\beta_3$).

Al aplicar la conocida fórmula de MCO a (10.2.3) obtenemos

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2} \tag{10.2.5}$$

Por consiguiente, aunque se puede estimar $\alpha$ en forma única, no hay forma de estimar $\beta_2$ y $\beta_3$ en forma igualmente única; matemáticamente,

$$\hat{\alpha} = \hat{\beta}_2 + \lambda\hat{\beta}_3 \tag{10.2.6}$$

nos proporciona una sola ecuación con dos incógnitas (observe que $\lambda$ está dada) y existen infinidad de soluciones para (10.2.6) con valores dados de $\hat{\alpha}$ y $\lambda$. Para expresar esto en términos concretos, sea $\hat{\alpha} = 0.8$ y $\lambda = 2$. Entonces:

$$0.8 = \hat{\beta}_2 + 2\hat{\beta}_3 \tag{10.2.7}$$

o

$$\hat{\beta}_2 = 0.8 - 2\hat{\beta}_3 \tag{10.2.8}$$

Ahora seleccione un valor de $\hat{\beta}_3$ arbitrariamente y tendrá una solución para $\hat{\beta}_2$. Seleccione otro valor para $\hat{\beta}_3$ y tendrá otra solución para $\hat{\beta}_2$. No importa cuánto lo intente, no existe un valor único para $\hat{\beta}_2$.

La conclusión del análisis anterior es que, en el caso de multicolinealidad perfecta, no puede obtenerse una solución única para los coeficientes de regresión individual. Pero observe que se puede obtener una solución única para combinaciones lineales de estos coeficientes. La combinación lineal $(\beta_2 + \lambda\beta_3)$ se estima en forma única con $\alpha$, dado el valor de $\lambda$.[9]

Asimismo, observe que en el caso de multicolinealidad perfecta, las varianzas y los errores estándar de $\hat{\beta}_2$ y $\hat{\beta}_3$ individualmente son infinitos. (Véase el ejercicio 10.21.)

## 10.3 Estimación en presencia de multicolinealidad "alta" pero "imperfecta"

La situación de multicolinealidad perfecta es un extremo patológico. Por lo general no existe una relación lineal exacta entre las variables $X$, en especial en información económica relacionada con series de tiempo. Por tanto, de regreso al modelo de tres variables en forma de desviación dado en (10.2.1), en lugar de multicolinealidad exacta podemos tener

$$x_{3i} = \lambda x_{2i} + v_i \tag{10.3.1}$$

donde $\lambda \neq 0$ y donde $v_i$ es un término de error estocástico tal que $\sum x_{2i} v_i = 0$. (¿Por qué?)

A propósito, los diagramas de Ballentine que aparecen en la figura 10.1*b*) a 10.1*e*) representan casos de colinealidad imperfecta.

En este caso, sería posible la estimación de los coeficientes de regresión $\beta_2$ y $\beta_3$. Por ejemplo, al sustituir (10.3.1) en (7.4.7) obtenemos

$$\hat{\beta}_2 = \frac{\sum(y_i x_{2i})\left(\lambda^2 \sum x_{2i}^2 + \sum v_i^2\right) - \left(\lambda \sum y_i x_{2i} + \sum y_i v_i\right)\left(\lambda \sum x_{2i}^2\right)}{\sum x_{2i}^2\left(\lambda^2 \sum x_{2i}^2 + \sum v_i^2\right) - \left(\lambda \sum x_{2i}^2\right)^2} \tag{10.3.2}$$

donde se aprovecha que $\sum x_{2i} v_i = 0$. Se deriva una expresión similar para $\hat{\beta}_3$.

[9] En teoría econométrica, una función del tipo $(\beta_2 + \lambda\beta_3)$ se conoce como **función estimable**.

Ahora, a diferencia de (10.2.2), no hay razón *a priori* para pensar que (10.3.2) no pueda estimarse. Desde luego, si $v_i$ es lo bastante pequeño, es decir, muy cercano a cero, (10.3.1) indicará colinealidad casi perfecta, y regresaremos al caso indeterminado de (10.2.2).

## 10.4 Multicolinealidad: ¿tanto para nada? Consecuencias teóricas de la multicolinealidad

Recuerde que si se satisfacen los supuestos del modelo clásico, los estimadores de MCO de los coeficientes de regresión son MELI (o MEI, si se añade el supuesto de normalidad). Ahora puede demostrarse que, aunque la multicolinealidad sea muy alta, como en el caso de *casi multicolinealidad*, los estimadores de MCO conservarán la propiedad MELI.[10] Entonces, ¿cuáles son los inconvenientes de la multicolinealidad? Christopher Achen comenta al respecto (tenga en cuenta también la cita de Leamer del principio de este capítulo):

> Los novatos en el estudio de la metodología en ocasiones se preocupan porque sus variables independientes estén correlacionadas: el llamado problema de multicolinealidad. Sin embargo, la multicolinealidad no viola los supuestos básicos de la regresión. Se presentarán estimaciones consistentes e insesgadas y sus errores estándar se estimarán en la forma correcta. El único efecto de la multicolinealidad tiene que ver con la dificultad de obtener los coeficientes estimados con errores estándar pequeños. Sin embargo, se presenta el mismo problema al contar con un número reducido de observaciones o al tener variables independientes con varianzas pequeñas. (De hecho, en el nivel teórico, los conceptos de multicolinealidad, número reducido de observaciones y varianzas pequeñas en las variables independientes forman parte esencial del mismo problema.) Por tanto, la pregunta "¿qué debe hacerse entonces con la multicolinealidad?" es similar a "¿qué debe hacerse si no se tienen muchas observaciones?" Al respecto no hay una respuesta estadística.[11]

Para referirse a la importancia del tamaño de la muestra, Goldberger acuñó el término **micronumerosidad**, como contraparte del exótico nombre polisílabo de multicolinealidad. De acuerdo con Goldberger, la **micronumerosidad exacta** (la contraparte de multicolinealidad exacta) surge cuando $n$, el tamaño de la muestra, es cero, en cuyo caso es imposible cualquier clase de estimación. La *casi micronumerosidad*, igual que la casi multicolinealidad, surge cuando el número de observaciones escasamente excede al número de parámetros que se va a estimar.

Leamer, Achen y Goldberger están en lo correcto al lamentar la falta de atención al problema del tamaño de la muestra, lo mismo que al problema de multicolinealidad. Por desgracia, en el trabajo aplicado que comprende información secundaria (es decir, información recopilada por alguna institución, como la información del PNB recopilada por el gobierno), es posible que un investigador por sí solo no pueda hacer gran cosa sobre el tamaño de la información muestral, y quizá deba enfrentar "la estimación de problemas lo bastante importantes para justificar su tratamiento [por ejemplo, la multicolinealidad] como una violación del modelo CRL [clásico de regresión lineal]".[12]

Primero, es cierto que aun en el caso de casi multicolinealidad los estimadores de MCO son insesgados. Pero el insesgamiento es una propiedad multimuestral o de muestreo repetido. Esto significa que, si mantenemos fijos los valores de $X$, si obtenemos muestras repetidas y calculamos los estimadores de MCO para cada una de esas muestras, el promedio de los valores muestrales se aproximará a los verdaderos valores poblacionales de los estimadores a medida que aumenta el número de las muestras. Pero esto nada dice sobre las propiedades de los estimadores en una muestra dada.

---

[10] Como la casi multicolinealidad por sí misma no viola los demás supuestos enumerados en el capítulo 7, los estimadores de MCO son MELI, como allí se indica.

[11] Christopher H. Achen, *Interpreting and Using Regression*, Sage, Beverly Hills, California, 1982, pp. 82-83.

[12] Peter Kennedy, *A Guide to Econometrics*, 3a. ed., The MIT Press, Cambridge, Massachusetts, 1992, p. 177.

Segundo, también es cierto que la colinealidad no destruye la propiedad de varianza mínima: en la clase de los estimadores lineales insesgados, los estimadores de MCO tienen varianza mínima; es decir, son eficientes. Pero esto no significa que la varianza de un estimador de MCO necesariamente sea pequeña (en relación con el valor del estimador) en cualquier muestra dada, como demostraremos en breve.

Tercero, la *multicolinealidad es en esencia un fenómeno (de regresión)* muestral en el sentido en que, aunque las variables $X$ no estén linealmente relacionadas en la población, pueden estarlo en la muestra particular disponible: cuando se postula la función de regresión teórica o poblacional (FRP), se considera que todas las variables $X$ incluidas del modelo ejercen una influencia separada o independiente sobre la variable dependiente $Y$. Pero puede suceder que en cualquier muestra dada con que se pruebe la FRP, alguna o todas las variables $X$ sean tan colineales que no sea posible aislar su influencia individual sobre $Y$. Es decir, la muestra falla aunque la teoría establezca que todas las $X$ son importantes. En resumen, la muestra puede no ser lo bastante "rica" para acomodar todas las variables $X$ en el análisis.

A manera de ilustración, reconsidere el ejemplo consumo-ingreso del capítulo 3. Los economistas teorizan que, además del ingreso, la riqueza del consumidor es también un determinante importante del gasto de consumo. Así, podemos escribir

$$\text{Consumo}_i = \beta_1 + \beta_2\,\text{Ingreso}_i + \beta_3\,\text{Riqueza}_i + u_i$$

Ahora, puede suceder que cuando se obtiene información sobre el ingreso y la riqueza, las dos variables pueden estar muy correlacionadas, aunque no en forma perfecta: la gente con mayor riqueza por lo general tiende a percibir mayores ingresos. Así, aunque, en teoría, el ingreso y la riqueza son candidatos lógicos para explicar el comportamiento del gasto de consumo, en la práctica (es decir, en la muestra) puede ser difícil distinguir las influencias separadas del ingreso y de la riqueza sobre el gasto de consumo.

Lo ideal para evaluar los efectos individuales de la riqueza y del ingreso sobre el gasto de consumo es un número suficiente de observaciones muestrales de individuos con riqueza pero con ingresos bajos, e individuos de altos ingresos con escasa riqueza (recuerde el supuesto 7). Aunque esto puede ser posible en los estudios de corte transversal (al incrementar el tamaño de la muestra), es muy difícil en el trabajo de series de tiempo agregadas.

Por todas estas razones, el hecho de que los estimadores de MCO sean MELI a pesar de la presencia de multicolinealidad es poco consuelo en la práctica. Se debe ver lo que sucede o puede suceder en una muestra dada, tema analizado en la siguiente sección.

## 10.5 Consecuencias prácticas de la multicolinealidad

En los casos de casi o alta multicolinealidad es probable que se presenten las siguientes consecuencias:

1. Aunque los estimadores de MCO son MELI, presentan varianzas y covarianzas grandes que dificultan la estimación precisa.
2. Debido a la consecuencia 1, los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de la "hipótesis nula cero" (es decir, que el verdadero coeficiente poblacional es cero).
3. También debido a la consecuencia 1, la razón $t$ de uno o más coeficientes tiende a ser estadísticamente no significativa.
4. Aunque la razón $t$ de uno o más coeficientes sea estadísticamente no significativa, $R^2$, la medida global de bondad de ajuste, puede ser muy alta.
5. Los estimadores de MCO y sus errores estándar son sensibles a pequeños cambios en los datos.

Las consecuencias anteriores se demuestran de la siguiente manera.

## Estimadores de MCO con varianzas y covarianzas grandes

Para ver varianzas y covarianzas grandes, recuerde que, para el modelo (10.2.1), las varianzas y covarianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$ están dadas por

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \tag{7.4.12}$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2\left(1 - r_{23}^2\right)} \tag{7.4.15}$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}} \tag{7.4.17}$$

donde $r_{23}$ es el coeficiente de correlación entre $X_2$ y $X_3$.

De (7.4.12) y (7.4.15) se desprende que, a medida que $r_{23}$ tiende a 1, es decir, a medida que aumenta la colinealidad, también lo hacen las varianzas de los dos estimadores y, en el límite, cuando $r_{23} = 1$, son infinitas. Es igualmente claro de (7.4.17) que, a medida que $r_{23}$ aumenta hacia 1, la covarianza de los dos estimadores también aumenta en valor absoluto. [*Nota:* cov ($\hat{\beta}_2$, $\hat{\beta}_3$) $\equiv$ cov ($\hat{\beta}_3$, $\hat{\beta}_2$).]

La velocidad con que se incrementan las varianzas y covarianzas se ve con el **factor inflacionario de la varianza** (**FIV**), que se define como

$$\text{FIV} = \frac{1}{\left(1 - r_{23}^2\right)} \tag{10.5.1}$$

El FIV muestra la forma como la varianza de un estimador se *infla* por la presencia de la multicolinealidad. A medida que $r_{23}^2$ se acerca a 1, el FIV se acerca a infinito. Es decir, a medida que el grado de colinealidad aumenta, la varianza de un estimador también y, en el límite, se vuelve infinita. Como se aprecia, si no hay colinealidad entre $X_2$ y $X_3$, el FIV será 1.

Con esta definición, (7.4.12) y (7.4.15) se expresan como

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2}\text{FIV} \tag{10.5.2}$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2}\text{FIV} \tag{10.5.3}$$

lo cual muestra que las varianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$ son directamente proporcionales al FIV.

Para dar alguna idea de la rapidez con que aumentan estas varianzas y covarianzas a medida que lo hace $r_{23}$, considere la tabla 10.1, que da estas varianzas y covarianzas para valores seleccionados de $r_{23}$. Como lo indica esta tabla, los aumentos en $r_{23}$ tienen un efecto drástico sobre las varianzas y covarianzas estimadas de los estimadores de MCO. Cuando $r_{23} = 0.50$, la var ($\hat{\beta}_2$) es 1.33 veces la varianza cuando $r_{23}$ es cero, pero, para cuando $r_{23}$ alcance 0.95, será alrededor de 10 veces más alta que cuando no hay colinealidad. Observe bien que un incremento de $r_{23}$ de 0.95 a 0.995 hace que la varianza estimada sea 100 veces la obtenida cuando la colinealidad es cero. Se observa el mismo efecto espectacular sobre la covarianza estimada. Todo esto se ve en la figura 10.2.

Los resultados recién analizados se extienden fácilmente al modelo con $k$ variables. En un modelo así, la varianza del $k$-ésimo coeficiente, como vimos en (7.5.6), se expresa como:

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1 - R_j^2}\right) \tag{7.5.6}$$

**TABLA 10.1**
**Efecto de incrementar $r_{23}$ sobre la var ($\hat\beta_2$) y la cov ($\hat\beta_2$, $\hat\beta_3$)**

| Valor de $r_{23}$ (1) | FIV (2) | var ($\hat\beta_2$) (3)* | $\frac{\text{var}(\hat\beta_2)(r_{23} \neq 0)}{\text{var}(\hat\beta_2)(r_{23} = 0)}$ (4) | cov ($\hat\beta_2$, $\hat\beta_3$) (5) |
|---|---|---|---|---|
| 0.00 | 1.00 | $\frac{\sigma^2}{\sum x_{2i}^2} = A$ | — | 0 |
| 0.50 | 1.33 | 1.33 × A | 1.33 | 0.67 × B |
| 0.70 | 1.96 | 1.96 × A | 1.96 | 1.37 × B |
| 0.80 | 2.78 | 2.78 × A | 2.78 | 2.22 × B |
| 0.90 | 5.76 | 5.26 × A | 5.26 | 4.73 × B |
| 0.95 | 10.26 | 10.26 × A | 10.26 | 9.74 × B |
| 0.97 | 16.92 | 16.92 × A | 16.92 | 16.41 × B |
| 0.99 | 50.25 | 50.25 × A | 50.25 | 49.75 × B |
| 0.995 | 100.00 | 100.00 × A | 100.00 | 99.50 × B |
| 0.999 | 500.00 | 500.00 × A | 500.00 | 499.50 × B |

*Nota:* $A = \frac{\sigma^2}{\sum x_{2i}^2}$

$B = \frac{-\sigma^2}{\sqrt{\sum x_{2i}^2 \sum x_{3i}^3}}$

× = multiplicación

*Para calcular el efecto de incrementar $r_{23}$ sobre var($\hat\beta_3$), observe que $A = \sigma^2 / \sum x_{3i}^2$ cuando $r_{23} = 0$; no obstante, los factores magnificadores de la varianza y covarianza permanecen iguales.

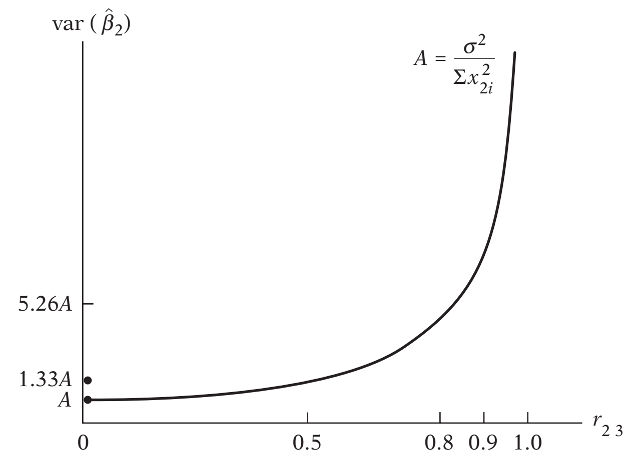



**FIGURA 10.2**
Comportamiento de la var ($\hat\beta_2$) como función de $r_{23}$.

donde $\hat\beta_j$ = coeficiente parcial (estimado) de la regresora $X_j$

$R_j^2 = R^2$ en la regresión de $X_j$ sobre las regresiones restantes $(k - 2)$ [*Nota:* Existen $(k - 1)$ regresoras en el modelo de regresión de $k$ variables.]

$\sum x_j^2 = \sum(X_j - \bar X_j)^2$

(7.5.6) también se expresa como

$$\text{var}(\hat\beta_j) = \frac{\sigma^2}{\sum x_j^2}\text{FIV}_j \tag{10.5.4}$$

Como puede observar en esta expresión, var ($\hat\beta_j$) es proporcional a $\sigma^2$ y a FIV, pero inversamente proporcional a $\sum x_j^2$. En consecuencia, el que var ($\hat\beta_j$) sea grande o pequeña depende de tres

**TABLA 10.2**
**Efecto de incrementar la colinealidad sobre el intervalo de confianza a 95% para $\beta_2$: $\hat{\beta}_2 \pm$ 1.96 ee ($\hat{\beta}_2$)**

| Valor de $r_{23}$ | Intervalo de confianza a 95% para $\beta_2$ |
|---|---|
| 0.00 | $\hat{\beta}_2 \pm 1.96\sqrt{\dfrac{\sigma^2}{\sum x_{2i}^2}}$ |
| 0.50 | $\hat{\beta}_2 \pm 1.96\sqrt{(1.33)}\sqrt{\dfrac{\sigma^2}{\sum x_{2i}^2}}$ |
| 0.95 | $\hat{\beta}_2 \pm 1.96\sqrt{(10.26)}\sqrt{\dfrac{\sigma^2}{\sum x_{2i}^2}}$ |
| 0.995 | $\hat{\beta}_2 \pm 1.96\sqrt{(100)}\sqrt{\dfrac{\sigma^2}{\sum x_{2i}^2}}$ |
| 0.999 | $\hat{\beta}_2 \pm 1.96\sqrt{(500)}\sqrt{\dfrac{\sigma^2}{\sum x_{2i}^2}}$ |

*Nota:* Se usa la distribución normal porque suponemos que se conoce $\sigma^2$ por conveniencia. De ahí que utilicemos 1.96, el factor de confianza de 95% para la distribución normal.

Los errores estándar que corresponden a los diversos valores de $r_{23}$ se obtienen de la tabla 10.1.

ingredientes: 1) $\sigma^2$, 2) FIV y 3) $\sum x_j^2$. Este último ingrediente, que se vincula al supuesto 8 del modelo clásico, establece que mientras más grande sea la variabilidad en una regresora, más pequeña será la varianza del coeficiente de esa regresora, si suponemos que los otros dos ingredientes son constantes; por tanto, será mayor la precisión para estimar dicho coeficiente.

Antes de seguir adelante, cabe observar que el inverso del FIV se conoce como **tolerancia** (TOL). Es decir,

$$\text{TOL}_j = \frac{1}{\text{FIV}_j} = \left(1 - R_j^2\right) \tag{10.5.5}$$

Cuando $R_j^2 = 1$ (es decir, colinealidad perfecta), $\text{TOL}_j = 0$, y cuando $R_j^2 = 0$ (es decir, no existe ninguna colinealidad), $\text{TOL}_j$ es 1. Debido a la estrecha conexión entre el FIV y la TOL pueden utilizarse de manera indistinta.

## Intervalos de confianza más amplios

Debido a los errores estándar grandes, los intervalos de confianza para los parámetros poblacionales relevantes tienden a ser mayores, como se ve en la tabla 10.2. Por ejemplo, cuando $r_{23} = 0.95$, el intervalo de confianza para $\beta_2$ es más grande que cuando $r_{23} = 0$ por un factor de $\sqrt{10.26}$, o alrededor de 3.

Por consiguiente, en casos de alta multicolinealidad, los datos muestrales pueden ser compatibles con un diverso conjunto de hipótesis. De ahí que aumente la probabilidad de aceptar una hipótesis falsa (es decir, un error tipo II).

## Razones *t* "no significativas"

Recuerde que para probar la hipótesis nula de que, por ejemplo, $\beta_2 = 0$, utilizamos la razón *t*, es decir, $\hat{\beta}_2/\text{ee}(\hat{\beta}_2)$ y comparamos el valor *t* estimado con el valor *t* crítico de la tabla *t*. Pero, como vimos, en casos de alta colinealidad los errores estándar estimados aumentan drásticamente, lo que disminuye los valores *t*. Por consiguiente, en tales casos se acepta cada vez con mayor facilidad la hipótesis nula de que el verdadero valor poblacional relevante es cero.[13]

[13] En términos de intervalos de confianza, a medida que aumenta el grado de colinealidad, el valor $\beta_2 = 0$ se ubica cada vez más en la región de aceptación.

## Una $R^2$ alta pero pocas razones *t* significativas

Considere el modelo de regresión lineal con *k* variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

En casos de alta colinealidad es posible encontrar, como acabamos de mencionar, que uno o más coeficientes parciales de pendiente son, de manera individual, no significativos estadísticamente con base en la prueba *t*. Aun así, $R^2$ en tales situaciones puede ser tan alto, digamos, superior a 0.9, que, con base en la prueba *F*, es posible rechazar convincentemente la hipótesis de que $\beta_2 = \beta_3 = \cdots = \beta_k = 0$. En realidad, ésta es una de las señales de multicolinealidad: valores *t* no significativos pero un $R^2$ global alto (y un valor *F* significativo).

Demostraremos lo anterior en la siguiente sección, pero este resultado no debe sorprender si tomamos en cuenta el análisis de las pruebas individuales comparadas con las pruebas conjuntas del capítulo 8. Como recordará, el problema real aquí consiste en que las covarianzas entre los estimadores, como indica la fórmula (7.4.17), están relacionadas con las correlaciones entre las regresoras.

## Sensibilidad de los estimadores de MCO y sus errores estándar ante cambios pequeños en los datos

Siempre que la multicolinealidad no sea perfecta, es posible la estimación de los coeficientes de regresión; sin embargo, las estimaciones y sus errores estándar se tornan muy sensibles aun al más ligero cambio de los datos.

Para ver esto considere la tabla 10.3. Con base en estos datos obtenemos la siguiente regresión múltiple:

$$\begin{aligned} \hat{Y}_i &= 1.1939 + 0.4463X_{2i} + 0.0030X_{3i} \\ &\quad (0.7737) \quad (0.1848) \quad (0.0851) \\ t &= (1.5431) \quad (2.4151) \quad (0.0358) \\ &\qquad R^2 = 0.8101 \qquad r_{23} = 0.5523 \\ &\qquad \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.00868 \qquad \text{gl} = 2 \end{aligned} \tag{10.5.6}$$

La regresión (10.5.6) muestra que ninguno de los coeficientes de regresión es individualmente significativo en los niveles de significancia convencionales de 1 o de 5%, a pesar de que $\hat{\beta}_2$ sea significativo en el nivel de 10% con base en la prueba *t* de una cola.

Ahora considere la tabla 10.4. La única diferencia entre las tablas 10.3 y 10.4 es que se intercambiaron el tercer y el cuarto valores de $X_3$. Con la información de la tabla 10.4 ahora obtenemos

$$\begin{aligned} \hat{Y}_i &= 1.2108 + 0.4014X_{2i} + 0.0270X_{3i} \\ &\quad (0.7480) \quad (0.2721) \quad (0.1252) \\ t &= (1.6187) \quad (1.4752) \quad (0.2158) \\ &\qquad R^2 = 0.8143 \qquad r_{23} = 0.8285 \\ &\qquad \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.0282 \qquad \text{gl} = 2 \end{aligned} \tag{10.5.7}$$

Como resultado de un ligero cambio en los datos vemos que $\hat{\beta}_2$, antes estadísticamente significativo en un nivel de significancia de 10%, deja ahora de serlo aun en ese nivel. Observe también que en (10.5.6) la cov $(\hat{\beta}_2, \hat{\beta}_3) = -0.00868$ mientras que en (10.5.7) es $-0.0282$, un aumento superior a tres veces su valor inicial. Todos estos cambios pueden atribuirse a un aumento de la multicolinealidad: en (10.5.6), $r_{23} = 0.5523$, mientras que en (10.5.7) este coeficiente es

**TABLA 10.3 Datos hipotéticos de $Y, X_2$ y $X_3$**

| $Y$ | $X_2$ | $X_3$ |
|---|---|---|
| 1 | 2 | 4 |
| 2 | 0 | 2 |
| 3 | 4 | 12 |
| 4 | 6 | 0 |
| 5 | 8 | 16 |

**TABLA 10.4 Datos hipotéticos de $Y, X_2$ y $X_3$**

| $Y$ | $X_2$ | $X_3$ |
|---|---|---|
| 1 | 2 | 4 |
| 2 | 0 | 2 |
| 3 | 4 | 0 |
| 4 | 6 | 12 |
| 5 | 8 | 16 |

de 0.8285. En forma similar, los errores estándar de $\hat{\beta}_2$ y $\hat{\beta}_3$ aumentan entre las dos regresiones, síntoma característico de la colinealidad.

Ya mencionamos que, en presencia de una alta colinealidad, no se pueden estimar los coeficientes de regresión individuales en forma precisa, pero que las combinaciones lineales de estos coeficientes se estiman con mayor exactitud. Esto se confirma con las regresiones (10.5.6) y (10.5.7). En la primera regresión, la suma de los dos coeficientes parciales de las pendientes es 0.4493, en tanto que en la segunda regresión dicha suma es 0.4284, prácticamente la misma. No sólo eso: sus errores estándar son prácticamente los mismos, 0.1550 frente a 0.1823.[14] Observe, sin embargo, que el coeficiente de $X_3$ cambió en forma notoria, de 0.003 a 0.027.

### Consecuencias de la micronumerosidad

En una parodia de las consecuencias de multicolinealidad y de manera informal, Goldberger cita consecuencias exactamente iguales del análisis basado en muestras pequeñas, es decir, de la micronumerosidad.[15] Se aconseja al lector consultar el análisis de Goldberger para ver la razón por la cual da la misma importancia a la micronumerosidad que a la multicolinealidad.

## 10.6 Ejemplo ilustrativo

**EJEMPLO 10.1**
*Gasto de consumo en relación con el ingreso y la riqueza*

Para ilustrar los puntos mencionados hasta ahora, reconsideremos el ejemplo consumo-ingreso de la introducción. La tabla 10.5 contiene datos hipotéticos sobre consumo, ingreso y riqueza. Si suponemos que el gasto de consumo se relaciona linealmente con el ingreso y la riqueza, entonces, con base en la tabla 10.5, obtenemos la siguiente regresión:

$$\begin{aligned} \hat{Y}_i &= 24.7747 + 0.9415X_{2i} - 0.0424X_{3i} \\ &\quad\ (6.7525) \quad (0.8229) \quad\quad (0.0807) \\ t &= \ (3.6690) \quad (1.1442) \quad\quad (-0.5261) \\ &\quad R^2 = 0.9635 \quad \bar{R}^2 = 0.9531 \quad \text{gl} = 7 \end{aligned} \tag{10.6.1}$$

[14] Estos errores estándar se obtienen de la fórmula

$$\text{ee}(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2\,\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}$$

Observe que, al aumentar la colinealidad, también lo hacen las varianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$, pero pueden compensarse si existe una alta covarianza negativa entre ambas, como lo indican claramente los resultados.

[15] Goldberger, *op. cit.,* pp. 248-250.

**TABLA 10.5** **Datos hipotéticos de gasto de consumo $Y$, ingreso $X_2$ y riqueza $X_3$**

| $Y$, \$ | $X_2$, \$ | $X_3$, \$ |
|---|---|---|
| 70 | 80 | 810 |
| 65 | 100 | 1 009 |
| 90 | 120 | 1 273 |
| 95 | 140 | 1 425 |
| 110 | 160 | 1 633 |
| 115 | 180 | 1 876 |
| 120 | 200 | 2 052 |
| 140 | 220 | 2 201 |
| 155 | 240 | 2 435 |
| 150 | 260 | 2 686 |

**TABLA 10.6**
**Tabla ANOVA para el ejemplo de consumo, ingreso y riqueza**

| Origen de la variación | SC | gl | SPC |
|---|---|---|---|
| Debido a la regresión | 8 565.5541 | 2 | 4 282.7770 |
| Debido a residuos | 324.4459 | 7 | 46.3494 |

La regresión (10.6.1) muestra que el ingreso y la riqueza explican en conjunto alrededor de 96% de la variación en los gastos de consumo. A pesar de esto, ningún coeficiente de las pendientes es estadísticamente significativo de manera individual. Además, no sólo la variable riqueza es estadísticamente no significativa, sino que también tiene el signo incorrecto. *A priori*, se esperaría una relación positiva entre el consumo y la riqueza. A pesar de que $\hat{\beta}_2$ y $\hat{\beta}_3$ no son significativos individualmente en términos estadísticos, si se prueba la hipótesis de que $\beta_2 = \beta_3 = 0$ simultáneamente, esta hipótesis puede rechazarse, como lo demuestra la tabla 10.6. Según el supuesto usual obtenemos

$$F = \frac{4\,282.7770}{46.3494} = 92.4019 \tag{10.6.2}$$

Como es obvio, este valor $F$ es muy significativo.

Es interesante observar este resultado desde un punto de vista geométrico. (Véase la figura 10.3.) Con base en la regresión (10.6.1) se establecieron intervalos de confianza individuales a 95% de confianza para $\beta_2$ y $\beta_3$ según el procedimiento usual del capítulo 8. Como muestran estos intervalos, cada uno de ellos, en forma individual, incluye el valor de cero. Por tanto, *individualmente* podemos aceptar la hipótesis de que las dos pendientes parciales son cero. Pero cuando establecemos el intervalo de confianza conjunto para probar la hipótesis de que $\beta_2 = \beta_3 = 0$, esa hipótesis no puede aceptarse, pues el intervalo de confianza conjunto, en realidad una elipse, no incluye el origen.[16] Como ya señalamos, cuando la colinealidad es alta, no son confiables las pruebas sobre las regresoras individuales; en tales casos, la prueba $F$ global es la que mostrará si $Y$ está relacionada con las diversas regresoras.

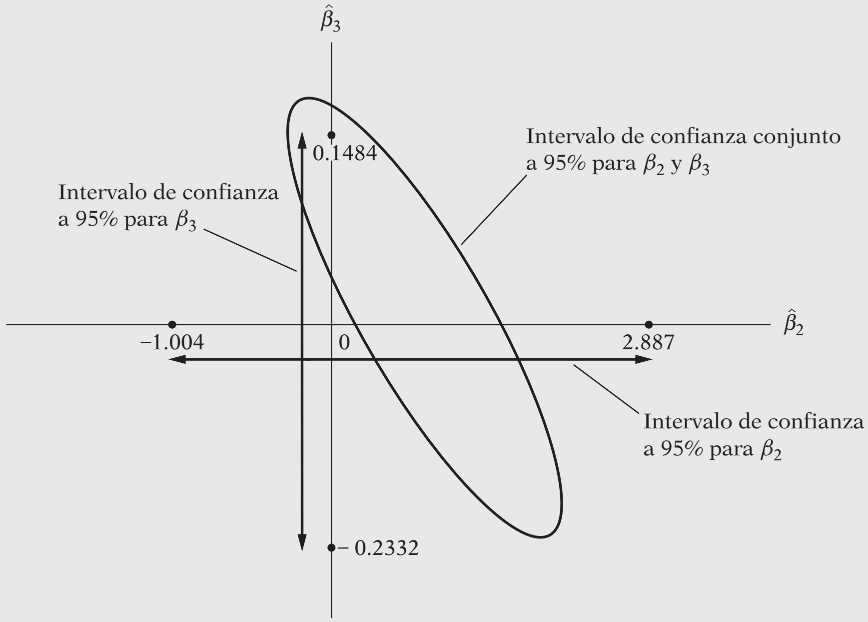
El ejemplo muestra en forma muy evidente lo que hace la multicolinealidad. El hecho de que la prueba $F$ sea significativa pero los valores $t$ de $X_2$ y $X_3$ no sean significativos individualmente implica que las dos variables están tan correlacionadas que es imposible aislar el impacto indi-

(*continúa*)

[16] Como mencionamos en la sección 5.3, el tema de intervalos de confianza conjuntos es muy complicado. El lector interesado puede consultar la referencia citada allí.

**EJEMPLO 10.1**
(*continuación*)

FIGURA 10.3 Intervalos de confianza individuales para $\beta_2$ y $\beta_3$ e intervalo de confianza conjunto (elipse) para $\beta_2$ y $\beta_3$.




vidual del ingreso o de la riqueza sobre el consumo. De hecho, si efectuamos la regresión de $X_3$ sobre $X_2$, obtenemos

$$\begin{aligned} \hat{X}_{3i} = &\ 7.5454 + 10.1909X_{2i} \\ &(29.4758) \quad (0.1643) \\ t = &\ (0.2560) \quad (62.0405) \qquad R^2 = 0.9979 \end{aligned} \tag{10.6.3}$$

lo cual muestra una colinealidad casi perfecta entre $X_3$ y $X_2$.

Ahora veamos lo que sucede si sólo efectuamos la regresión de $Y$ sobre $X_2$:

$$\begin{aligned} \hat{Y}_i = &\ 24.4545 + 0.5091X_{2i} \\ &(6.4138) \quad (0.0357) \\ t = &\ (3.8128) \quad (14.2432) \qquad R^2 = 0.9621 \end{aligned} \tag{10.6.4}$$

En (10.6.1), la variable ingreso no era estadísticamente significativa, mientras que ahora es muy significativa. Si en lugar de efectuar la regresión de $Y$ sobre $X_2$ lo hacemos sobre $X_3$, obtenemos

$$\begin{aligned} \hat{Y}_i = &\ 24.411 + 0.0498X_{3i} \\ &(6.874) \quad (0.0037) \\ t = &\ (3.551) \quad (13.29) \qquad R^2 = 0.9567 \end{aligned} \tag{10.6.5}$$

Se observa que la riqueza tiene ahora un impacto significativo sobre el gasto de consumo, mientras que en (10.6.1) no tenía ninguno.

Las regresiones (10.6.4) y (10.6.5) muestran con toda claridad que, en situaciones de extrema multicolinealidad, eliminar la variable altamente colineal con frecuencia provoca que la otra variable $X$ se torne estadísticamente significativa. Este resultado sugiere que una forma de evadir la colinealidad extrema es eliminar la variable colineal, tema que veremos con mayor detalle en la sección 10.8.

**EJEMPLO 10.2**
*Función de consumo para Estados Unidos, 1947-2000*

Fuente: Véase la tabla 7.12.

A continuación consideraremos un grupo concreto de datos sobre gasto de consumo real (C), ingreso personal disponible real (Yd), riqueza real (W) y tasa de interés real (I) para Estados Unidos de 1947 a 2000. Los datos originales se presentan en la tabla 10.7.

**TABLA 10.7** **Gasto de consumo en Estados Unidos del periodo 1947-2000**

| Año | C | Yd | W | I |
|---|---|---|---|---|
| 1947 | 976.4 | 1 035.2 | 5 166.815 | −10.35094 |
| 1948 | 998.1 | 1 090 | 5 280.757 | −4.719804 |
| 1949 | 1 025.3 | 1 095.6 | 5 607.351 | 1.044063 |
| 1950 | 1 090.9 | 1 192.7 | 5 759.515 | 0.407346 |
| 1951 | 1 107.1 | 1 227 | 6 086.056 | −5.283152 |
| 1952 | 1 142.4 | 1 266.8 | 6 243.864 | −0.277011 |
| 1953 | 1 197.2 | 1 327.5 | 6 355.613 | 0.561137 |
| 1954 | 1 221.9 | 1 344 | 6 797.027 | −0.138476 |
| 1955 | 1 310.4 | 1 433.8 | 7 172.242 | 0.261997 |
| 1956 | 1 348.8 | 1 502.3 | 7 375.18 | −0.736124 |
| 1957 | 1 381.8 | 1 539.5 | 7 315.286 | −0.260683 |
| 1958 | 1 393 | 1 553.7 | 7 869.975 | −0.57463 |
| 1959 | 1 470.7 | 1 623.8 | 8 188.054 | 2.295943 |
| 1960 | 1 510.8 | 1 664.8 | 8 351.757 | 1.511181 |
| 1961 | 1 541.2 | 1 720 | 8 971.872 | 1.296432 |
| 1962 | 1 617.3 | 1 803.5 | 9 091.545 | 1.395922 |
| 1963 | 1 684 | 1 871.5 | 9 436.097 | 2.057616 |
| 1964 | 1 784.8 | 2 006.9 | 10 003.4 | 2.026599 |
| 1965 | 1 897.6 | 2 131 | 10 562.81 | 2.111669 |
| 1966 | 2 006.1 | 2 244.6 | 10 522.04 | 2.020251 |
| 1967 | 2 066.2 | 2 340.5 | 11 312.07 | 1.212616 |
| 1968 | 2 184.2 | 2 448.2 | 12 145.41 | 1.054986 |
| 1969 | 2 264.8 | 2 524.3 | 11 672.25 | 1.732154 |
| 1970 | 2 317.5 | 2 630 | 11 650.04 | 1.166228 |
| 1971 | 2 405.2 | 2 745.3 | 12 312.92 | −0.712241 |
| 1972 | 2 550.5 | 2 874.3 | 13 499.92 | −0.155737 |
| 1973 | 2 675.9 | 3 072.3 | 13 080.96 | 1.413839 |
| 1974 | 2 653.7 | 3 051.9 | 11 868.79 | −1.042571 |
| 1975 | 2 710.9 | 3 108.5 | 12 634.36 | −3.533585 |
| 1976 | 2 868.9 | 3 243.5 | 13 456.78 | −0.656766 |
| 1977 | 2 992.1 | 3 360.7 | 13 786.31 | −1.190427 |
| 1978 | 3 124.7 | 3 527.5 | 14 450.5 | 0.113048 |
| 1979 | 3 203.2 | 3 628.6 | 15 340 | 1.70421 |
| 1980 | 3 193 | 3 658 | 15 964.95 | 2.298496 |
| 1981 | 3 236 | 3 741.1 | 15 964.99 | 4.703847 |
| 1982 | 3 275.5 | 3 791.7 | 16 312.51 | 4.449027 |
| 1983 | 3 454.3 | 3 906.9 | 16 944.85 | 4.690972 |
| 1984 | 3 640.6 | 4 207.6 | 17 526.75 | 5.848332 |
| 1985 | 3 820.9 | 4 347.8 | 19 068.35 | 4.330504 |
| 1986 | 3 981.2 | 4 486.6 | 20 530.04 | 3.768031 |
| 1987 | 4 113.4 | 4 582.5 | 21 235.69 | 2.819469 |
| 1988 | 4 279.5 | 4 784.1 | 22 331.99 | 3.287061 |

(*continúa*)

**EJEMPLO 10.2** (*continuación*)

TABLA 10.7 *Continuación*

| Año | C | Yd | W | I |
|---|---|---|---|---|
| 1989 | 4 393.7 | 4 906.5 | 23 659.8 | 4.317956 |
| 1990 | 4 474.5 | 5 014.2 | 23 105.13 | 3.595025 |
| 1991 | 4 466.6 | 5 033 | 24 050.21 | 1.802757 |
| 1992 | 4 594.5 | 5 189.3 | 24 418.2 | 1.007439 |
| 1993 | 4 748.9 | 5 261.3 | 25 092.33 | 0.62479 |
| 1994 | 4 928.1 | 5 397.2 | 25 218.6 | 2.206002 |
| 1995 | 5 075.6 | 5 539.1 | 27 439.73 | 3.333143 |
| 1996 | 5 237.5 | 5 677.7 | 29 448.19 | 3.083201 |
| 1997 | 5 423.9 | 5 854.5 | 32 664.07 | 3.12 |
| 1998 | 5 683.7 | 6 168.6 | 35 587.02 | 3.583909 |
| 1999 | 5 968.4 | 6 320 | 39 591.26 | 3.245271 |
| 2000 | 6 257.8 | 6 539.2 | 38 167.72 | 3.57597 |

Empleamos lo siguiente para el análisis:

$$\ln C_t = \beta_1 + \beta_2 \ln Yd_t + \beta_3 \ln W_t + \beta_4 I_t + u_t \qquad \textbf{(10.6.6)}$$

donde ln significa logaritmo.

En este modelo, los coeficientes $\beta_2$ y $\beta_3$ dan las elasticidades del ingreso y la riqueza, respectivamente (¿por qué?), y $\beta_4$ da la semielasticidad (¿por qué?). Los resultados de la regresión (10.6.6) se presentan en la siguiente tabla:

Variable dependiente: LOG (C)
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | -0.467711 | 0.042778 | -10.93343 | 0.0000 |
| LOG (YD) | 0.804873 | 0.017498 | 45.99836 | 0.0000 |
| LOG (RIQUEZA) | 0.201270 | 0.017593 | 11.44060 | 0.0000 |
| INTERÉS | -0.002689 | 0.000762 | -3.529265 | 0.0009 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.999560 | Media de la variable dependiente | 7.826093 |
| *R* cuadrada ajustada | 0.999533 | Desviación estándar de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.011934 | Criterio de información de Akaike | -5.947703 |
| Suma de cuadrados residual | 0.007121 | Criterio de Schwarz | -5.800371 |
| Log verosimilitud | 164.5880 | Criterio de Hannan-Quinn | -5.890883 |
| Estadístico *F* | 37 832.59 | Estadístico de Durbin-Watson | 1.289219 |
| Probabilidad (estadístico *F*) | 0.000000 | | |

*Nota:* LOG significa logaritmo natural.

Los resultados demuestran que todos los coeficientes estimados son muy significativos desde el punto de vista estadístico, pues sus valores $p$ son muy pequeños. Los coeficientes estimados se interpretan como sigue: la elasticidad del ingreso es $\approx$ 0.80, lo que indica que, cuando las demás variables se mantienen constantes, si el ingreso aumenta 1%, la media del gasto de con-

sumo aumenta alrededor de 0.8%. El coeficiente de riqueza es ≈ 0.20, lo que significa que si la riqueza aumenta 1%, la media del consumo se incrementa sólo 0.2%, de nuevo cuando las demás variables se mantienen constantes. El coeficiente de la variable tasa de interés indica que, a medida que la tasa de interés aumenta *un punto porcentual,* el gasto de consumo *disminuye* 0.26%, *ceteris paribus.*

Todas las regresoras tienen signos que concuerdan con las expectativas previas, es decir, el ingreso y la riqueza tienen efecto positivo en el consumo, pero la tasa de interés produce un efecto negativo.

¿Hay que preocuparse por el problema de la multicolinealidad en este caso? Al parecer no, porque todos los coeficientes tienen los signos correctos, cada coeficiente es muy significativo estadísticamente en lo individual y el valor $F$ también es estadísticamente muy significativo, lo que indica que, en conjunto, todas las variables tienen efecto significativo en el gasto de consumo. El valor $R^2$ también es muy alto.

Por supuesto, casi siempre existe cierto grado de colinealidad entre las variables económicas. Con tal de que no sea exacto se pueden estimar los parámetros del modelo. Por el momento, lo único que se puede decir es que, en el presente ejemplo, la colinealidad, si la hay, no parece muy marcada. Sin embargo, en la sección 10.7 presentamos algunas pruebas de diagnóstico para detectar la colinealidad y reexaminar la función de consumo de Estados Unidos para determinar si le afecta el problema de la colinealidad.

## 10.7 Detección de la multicolinealidad

Después de estudiar las características y las consecuencias de la multicolinealidad, el interrogante natural es: ¿cómo conocer la presencia de colinealidad en cualquier situación dada, en especial en modelos con más de dos variables explicativas? Aquí es útil la advertencia de Kmenta:

> **1.** La multicolinealidad es una cuestión de grado y no de clase. La distinción importante no es entre presencia o ausencia de multicolinealidad, sino entre sus diferentes grados.
>
> **2.** Como la multicolinealidad se refiere a la condición de las variables explicativas que son no estocásticas por supuestos, es una característica de la muestra y no de la población.
>
> Por consiguiente, no es necesario "llevar a cabo pruebas sobre multicolinealidad", pero, si se desea, es posible medir su grado en cualquier muestra determinada.[17]

Como la multicolinealidad es en esencia un fenómeno de tipo muestral que surge de información sobre todo no experimental recopilada en la mayoría de las ciencias sociales, no hay un método único para detectarla o medir su fuerza. Lo que se tiene en realidad son ciertas reglas prácticas, algunas informales y otras formales, pero todas reglas prácticas. Consideremos algunas de ellas.

1. **Una $R^2$ elevada pero pocas razones $t$ significativas.** Como ya mencionamos, es un síntoma "clásico" de multicolinealidad. Si $R^2$ es alta, es decir, está por encima de 0.8, la prueba $F$, en la mayoría de los casos, rechazará la hipótesis de que los coeficientes parciales de pendiente son simultáneamente iguales a cero, pero las pruebas $t$ individuales mostrarán que ningún coeficiente parcial de pendiente, o muy pocos, son estadísticamente diferentes de cero. Demostramos lo anterior con claridad en el ejemplo de consumo-ingreso-riqueza.

Aunque este diagnóstico es razonable, su desventaja es que "es demasiado fuerte, en el sentido de que la multicolinealidad se considera dañina únicamente cuando no se puede separar la totalidad de las influencias de las variables explicativas sobre $Y$".[18]

---

[17] Jan Kmenta, *Elements of Econometrics,* 2a. ed., Macmillan, Nueva York, 1986, p. 431.

[18] *Ibid.,* p. 439.

2. **Altas correlaciones entre parejas de regresoras.** Otra regla práctica recomendable consiste en observar el coeficiente de correlación de orden cero o entre dos regresoras. Si éste es alto, digamos, superior a 0.8, la multicolinealidad es un problema grave. La desventaja con este criterio es que, aunque las altas correlaciones de orden cero pueden sugerir la presencia de colinealidad, no es necesario que dichas correlaciones sean altas para tener colinealidad en un determinado caso específico. En términos un poco técnicos: *las correlaciones de orden cero elevadas son una condición suficiente pero no necesaria para la existencia de multicolinealidad, debido a que puede existir a pesar de que las correlaciones de orden cero o correlaciones simples sean comparativamente bajas* (es decir, inferiores a 0.50). Para apreciar esta relación, suponga un modelo con cuatro variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

y suponga que

$$X_{4i} = \lambda_2 X_{2i} + \lambda_3 X_{3i}$$

donde $\lambda_2$ y $\lambda_3$ son constantes, sin ser las dos iguales a cero. Obvio, $X_4$ es una combinación lineal exacta de $X_2$ y $X_3$, que da $R^2_{4.23} = 1$, el coeficiente de determinación en la regresión de $X_4$ sobre $X_2$ y $X_3$.

Ahora recordemos la fórmula (7.11.5) del capítulo 7 para escribir

$$R^2_{4.23} = \frac{r^2_{42} + r^2_{43} - 2r_{42}r_{43}r_{23}}{1 - r^2_{23}} \tag{10.7.1}$$

Pero, como $R^2_{4.23} = 1$ por la existencia de colinealidad perfecta, obtenemos

$$1 = \frac{r^2_{42} + r^2_{43} - 2r_{42}r_{43}r_{23}}{1 - r^2_{23}} \tag{10.7.2}$$

No es difícil ver que (10.7.2) se satisface con $r_{42} = 0.5$, $r_{43} = 0.5$ y $r_{23} = -0.5$, que no son valores muy altos.

Por consiguiente, en los modelos donde hay más de dos variables explicativas, la correlación simple o de orden cero no proporciona una guía infalible sobre la presencia de multicolinealidad. Claro que si sólo existen dos variables explicativas, bastarán las correlaciones de orden cero.

3. **Examen de las correlaciones parciales.** Debido al problema recién descrito, que se basa en correlaciones de orden cero, Farrar y Glauber sugieren que deben observarse, en lugar de ellas, los coeficientes de correlación parcial.[19] De esta forma, en la regresión de $Y$ sobre $X_2$, $X_3$ y $X_4$, si se encuentra que $R^2_{1.234}$ es muy elevada pero $r^2_{12.34}$, $r^2_{13.24}$ y $r^2_{14.23}$ son comparativamente bajas, esto puede sugerir que las variables $X_2$, $X_3$ y $X_4$ están muy intercorrelacionadas y que por lo menos una de estas variables es superflua.

Si bien puede ser útil un estudio de correlaciones parciales, nada garantiza que proporcionen una guía infalible sobre multicolinealidad, pues puede suceder que tanto $R^2$ como todas las correlaciones parciales sean lo bastante altas. Sin embargo, y tal vez más importante, C. Robert Wichers mostró[20] que la prueba de correlación parcial de Farrar-Glauber es ineficaz en el sentido

[19] D.E. Farrar y R.R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited", *Review of Economics and Statistics*, vol. 49, 1967, pp. 92-107.

[20] "The Detection of Multicollinearity: A Comment", *Review of Economics and Statistics*, vol. 57, 1975, pp. 365-366.

de que una determinada correlación parcial puede ser compatible con diferentes patrones de multicolinealidad. La prueba de Farrar-Glauber también recibió fuertes críticas de T. Krishna Kumar,[21] John O'Hagan y Brendan McCabe.[22]

4. **Regresiones auxiliares.** Como la multicolinealidad surge porque una o más de las regresoras son combinaciones lineales exactas o aproximadas de las demás regresoras, una forma de determinar cuál variable $X$ está relacionada con las demás variables $X$ es efectuar la regresión de cada $X_i$ sobre las variables $X$ restantes y calcular la $R^2$ correspondiente, que se designa $R_i^2$; cada una de estas regresiones se denomina **regresión auxiliar**, auxiliar a la regresión principal de $Y$ sobre las $X$. Así, conforme a la relación entre $F$ y $R^2$ establecida en (8.4.11), la variable

$$F_i = \frac{R^2_{x_i \cdot x_2 x_3 \cdots x_k} \big/ (k-2)}{\left(1 - R^2_{x_i \cdot x_2 x_3 \cdots x_k}\right) \big/ (n-k+1)} \tag{10.7.3}$$

sigue la distribución $F$ con $k-2$ y $n-k+1$ gl. En la ecuación (10.7.3), $n$ representa el tamaño de la muestra, $k$ representa el número de variables explicativas incluyendo el intercepto y $R^2_{x_i \cdot x_2 x_3 \cdots x_k}$ es el coeficiente de determinación en la regresión de la variable $X_i$ sobre las variables $X$ restantes.[23]

Si la $F$ calculada excede a la $F_i$ crítica en el nivel de significancia seleccionado, se dice que la $X_i$ particular es colineal con las demás $X$; si no excede a la $F_i$ crítica, se dice que ésta no es colineal con las demás $X$, en cuyo caso se puede mantener la variable en el modelo. Si $F_i$ es estadísticamente significativa, aún hay que decidir si la $X_i$ en consideración debe eliminarse del modelo. Analizaremos este aspecto con más detalle en la sección 10.8.

Sin embargo, este método no carece de desventajas, pues

> . . . si la multicolinealidad comprende sólo unas cuantas variables, de forma que las regresiones auxiliares no sufran de multicolinealidad extensa, los coeficientes estimados pueden revelar la naturaleza de la dependencia lineal entre las regresoras. Por desgracia, si existen diversas asociaciones lineales complejas, este ejercicio de ajuste de curva puede no tener gran valor, pues será difícil identificar las interrelaciones separadas.[24]

En lugar de probar formalmente todos los valores $R^2$ auxiliares, se puede adoptar la **regla práctica de Klein**, que sugiere que la multicolinealidad puede ser un problema complicado solamente si la $R^2$ obtenida de una regresión auxiliar es mayor que la $R^2$ global, es decir, si se obtiene de la regresión de $Y$ sobre todas las regresoras.[25] Por cierto, al igual que todas las demás reglas prácticas, ésta debe utilizarse con buen criterio.

5. **Valores propios e índice de condición.** Mediante *EViews* y *Stata* podemos calcular los *valores propios* y el *índice de condición* para diagnosticar la multicolinealidad. No analizaremos aquí el tema de los valores propios, pues implicaría abordar temas de álgebra matricial, fuera

---

[21] "Multicollinearity in Regression Analysis", *Review of Economics and Statistics*, vol. 57, 1975, pp. 366-368.

[22] "Tests for the Severity of Multicollinearity in Regression Analysis: A Comment", *Review of Economics and Statistics*, vol. 57, 1975, pp. 368-370.

[23] Por ejemplo, $R^2_{x_2}$ se obtiene mediante la regresión de $X_{2i}$ de la siguiente manera: $X_{2i} = a_1 + a_3 X_{3i} + a_4 X_{4i} + \cdots + a_k X_{ki} + \hat{u}_i$.

[24] George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lütkepohl y Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, Nueva York, 1982, p. 621.

[25] Lawrence R. Klein, *An Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1962, p. 101.

del alcance de este libro. Sin embargo, a partir de estos valores propios puede derivarse lo que se conoce como **número de condición *k***, definido como

$$k = \frac{\text{Valor propio máximo}}{\text{Valor propio mínimo}}$$

y el **índice de condición (IC)**, definido como

$$\text{IC} = \sqrt{\frac{\text{Valor propio máximo}}{\text{Valor propio mínimo}}} = \sqrt{k}$$

**Entonces tenemos esta regla práctica:** Si $k$ está entre l00 y 1 000, existe una multicolinealidad que va de moderada a fuerte, mientras que si excede de 1 000, existe multicolinealidad grave. De otro modo, si el IC ($= \sqrt{k}$) está entre 10 y 30, hay multicolinealidad entre moderada y fuerte, y si excede de 30, una multicolinealidad grave.

Para el ejemplo ilustrativo del apéndice 7A.5, el valor propio más pequeño es 3.786 y el valor propio más grande es 187.5269, por lo que $k = 187.5269/3.786$, o alrededor de 49.53. Por tanto, $\text{IC} = \sqrt{49.53} = 7.0377$. Tanto $k$ como IC indican que no existe un problema grave de colinealidad. Por cierto, observe que un valor propio bajo (en relación con el valor propio máximo) es, por lo general, indicativo de dependencias casi lineales en los datos.

Algunos autores consideran que e1 índice de condición es el mejor diagnóstico de multicolinealidad disponible. Sin embargo, esta opinión no es muy aceptada. Así, el IC es sólo una regla práctica, quizá un poco más compleja. Para mayores detalles, el lector puede consultar las referencias.[26]

6. **Tolerancia y factor de inflación de la varianza.** Ya vimos el FIV y la TOL. Conforme $R_j^2$ —el coeficiente de determinación en la regresión de la regresora $X_j$ sobre las regresoras restantes del modelo— se aproxima a la unidad, es decir, conforme se incrementa la colinealidad de $X_j$ con las demás regresoras, FIV también aumenta, y en el límite puede ser infinito.

Algunos autores utilizan, por consiguiente, el FIV como indicador de la multicolinealidad: entre mayor es el valor del $\text{FIV}_j$, mayor "problema" o colinealidad tiene la variable $X_j$. ¿Pero, cuánto debe ascender el FIV antes de que una regresora se convierta en un problema? **Como regla práctica**, si el FIV de una variable es superior a 10 (esto sucede si $R_j^2$ excede de 0.90), se dice que esa variable es muy colineal.[27]

Desde luego, puede utilizarse $\text{TOL}_j$ como medida de la multicolinealidad, en vista de su estrecha conexión con $\text{FIV}_j$. Mientras más cerca esté $\text{TOL}_j$ de cero, mayor será el grado de colinealidad de esa variable respecto de las demás regresoras. Por otra parte, mientras más cerca esté $\text{TOL}_j$ de 1, mayor será la evidencia de que $X_j$ no es colineal con las demás regresoras.

El FIV (o tolerancia) como medida de colinealidad no está libre de crítica. Como indica (10.5.4), $\text{var}(\hat{\beta}_j)$ depende de tres factores: $\sigma^2$, $\sum x_j^2$ y $\text{FIV}_j$. Un FIV alto se contrarresta por una $\sigma^2$ baja o una $\sum x_j^2$ alta. De otra forma: un FIV alto no es condición necesaria ni suficiente para obtener varianzas y errores estándar altos. Por consiguiente, la alta multicolinealidad, como la mide un FIV alto, puede no necesariamente ocasionar errores estándar altos. En todo este análisis, los términos *alto* y *bajo* son relativos.

7. **Diagrama de dispersión.** Es una buena práctica usar un diagrama de dispersión para ver cómo se relacionan las diversas variables de un modelo de regresión. La figura 10.4 presenta el

[26] Véase sobre todo D.A. Belsley, E. Kuh y R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, Nueva York, 1980, capítulo 3. Sin embargo, este libro no es para principiantes.

[27] Véase David G. Kleinbaum, Lawrence L. Kupper y Keith E. Muller, *Applied Regression Analysis and Other Multivariate Methods*, 2a. ed., PWS-Kent, Boston, Massachusetts, 1988, p. 210.

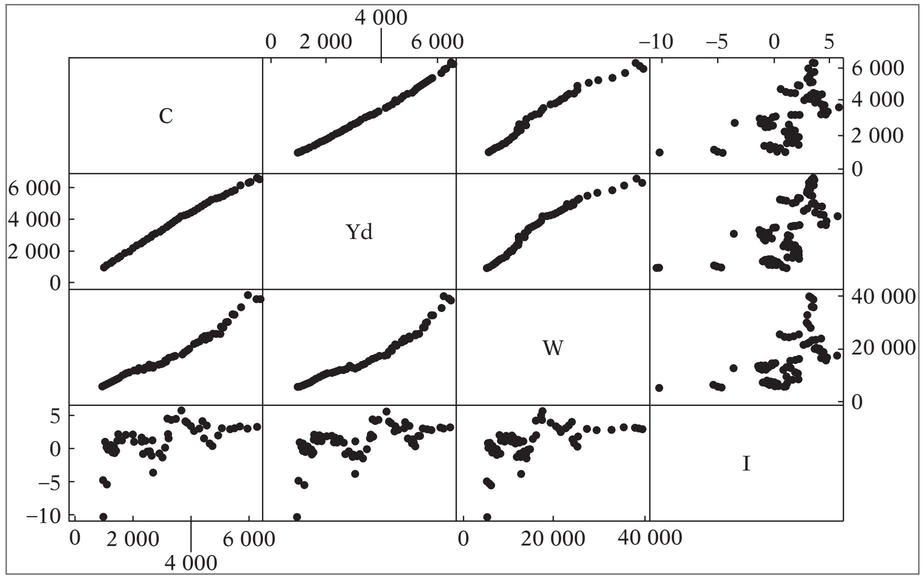



**FIGURA 10.4**
Diagrama de dispersión de los datos del ejemplo 10.2.

diagrama de dispersión del ejemplo de consumo analizado en la sección anterior (ejemplo 10.2). Se trata de un diagrama de cuatro por cuatro cuadros porque hay cuatro variables en el modelo, una variable dependiente (C) y tres variables explicativas: ingreso personal disponible real (Yd), riqueza real (W) y tasa de interés real (I).

Primero considere la diagonal principal, de la esquina superior izquierda a la esquina inferior derecha. No hay puntos de dispersión en estos cuadros en la diagonal principal. Si los hubiera, tendrían un coeficiente de correlación de 1, pues las gráficas serían de una variable dada sobre sí misma. Los cuadros fuera de la diagonal muestran intercorrelaciones entre las variables. Por ejemplo, el cuadro de riqueza (W) muestra que la riqueza y el ingreso están muy correlacionados (el coeficiente de correlación entre los dos es 0.97), pero no de manera perfecta. Si tuvieran correlación perfecta (es decir, si tuvieran un coeficiente de correlación de 1), no habríamos podido estimar la regresión (10.6.6) porque habría una relación lineal exacta entre riqueza e ingreso. El diagrama de dispersión también muestra que la tasa de interés no está muy correlacionada con las otras tres variables.

Como la función de diagrama de dispersión se incluye ahora en varios programas estadísticos, este diagnóstico debe tomarse en consideración junto con los que estudiamos antes. No obstante, hay que recordar que las correlaciones simples entre parejas de variables pueden no ser un indicador definitivo de colinealidad, como ya señalamos.

Para concluir la detección de la multicolinealidad, reiteramos que los diversos métodos son en esencia "expediciones de pesca", pues no puede decirse cuáles funcionan en una aplicación particular. Sin embargo, no se puede hacer mucho al respecto, pues la multicolinealidad es un problema específico de una muestra dada sobre la cual el investigador puede no tener mucho control, sobre todo si los datos son no experimentales por naturaleza, como es lo común para los investigadores de las ciencias sociales.

Nuevamente, como una parodia de multicolinealidad, Goldberger cita diversas formas de detectar la micronumerosidad, como el desarrollo de valores críticos del tamaño de la muestra, $n^*$, tales que la micronumerosidad es un problema sólo si el tamaño real de la muestra $n$ es más pequeño que $n^*$. Lo importante de la parodia de Goldberger es destacar que el tamaño pequeño de la muestra y la falta de variabilidad en las variables explicativas pueden ocasionar problemas por lo menos tan graves como los debidos a la multicolinealidad.

## 10.8 Medidas correctivas

¿Qué puede hacerse si la multicolinealidad es grave? Hay dos posibilidades: 1) no hacer nada o 2) seguir algunas reglas prácticas.

### No hacer nada

Blanchard expresa de la siguiente manera la corriente de pensamiento que aboga por "no hacer nada":[28]

> Cuando los estudiantes efectúan por primera vez la regresión de mínimos cuadrados ordinarios (MCO), el primer problema que suelen afrontar es el de la multicolinealidad. Muchos concluyen que hay algo malo con los MCO; otros recurren a nuevas y con frecuencia creativas técnicas a fin de darle la vuelta al problema. Pero eso está mal. La multicolinealidad es la voluntad de Dios, no un problema con los MCO ni con la técnica estadística en general.

Lo que Blanchard afirma es que la multicolinealidad es en esencia un problema de deficiencia de datos (de nuevo, micronumerosidad), y en algunas ocasiones no hay opción respecto de los datos disponibles para el análisis empírico.

Asimismo, no es que todos los coeficientes en un modelo de regresión sean estadísticamente insignificantes. Al contrario, aunque no se puedan estimar uno o más coeficientes de regresión con gran precisión, es posible calcular una combinación lineal de ellos (es decir, una función estimable) con relativa eficiencia. Como vimos en (10.2.3), $\alpha$ se calcula de forma única, aunque no puedan estimarse sus dos componentes dados ahí de manera individual. Algunas veces esto es lo mejor que se puede hacer con un determinado conjunto de datos.[29]

### Procedimientos de reglas prácticas

Se pueden intentar las siguientes reglas prácticas para abordar el problema de la multicolinealidad; el éxito depende de la gravedad de la multicolinealidad.

1. **Información *a priori*.** Suponga que consideramos el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

donde $Y$ = consumo, $X_2$ = ingreso y $X_3$ = riqueza. Como ya mencionamos, las variables ingreso y riqueza tienden a ser muy colineales. Pero suponga que, *a priori*, creemos que $\beta_3 = 0.10\beta_2$; es decir, la tasa de cambio del consumo respecto de la riqueza es una décima parte de la correspondiente respecto del ingreso. Podemos entonces efectuar la siguiente regresión:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + 0.10\, \beta_2 X_{3i} + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned}$$

donde $X_i = X_{2i} + 0.1X_{3i}$. Una vez obtenido $\hat{\beta}_2$ podemos estimar $\hat{\beta}_3$ a partir de la relación postulada entre $\beta_2$ y $\beta_3$.

¿Cómo obtener información *a priori*? Puede provenir de un trabajo empírico anterior, en donde el problema de colinealidad resultó ser menos grave o de la teoría relevante que soporta

---

[28] O.J. Blanchard, "Comment", *Journal of Business and Economic Statistics,* vol. 5, 1967, pp. 449-451. La cita se toma de Peter Kennedy, *A Guide to Econometrics,* 4a. ed., MIT Press, Cambridge, Massachusetts, 1998, p. 190.

[29] Hay un interesante análisis sobre este tema en J. Conlisk, "When Collinearity is Desirable", *Western Economic Journal*, vol. 9, 1971, pp. 393-407.

el campo de estudio. Por ejemplo, en la función de producción tipo Cobb-Douglas (7.9.1), si esperamos que prevalezcan los rendimientos constantes a escala, entonces $(\beta_2 + \beta_3) = 1$, en cuyo caso podemos efectuar la regresión (8.6.14), con la regresión de la razón producto-trabajo sobre la razón capital-trabajo. Si existe colinealidad entre el trabajo y el capital, como suele ser el caso en la mayor parte de la información muestral, dicha transformación puede reducir o eliminar el problema de colinealidad. Pero es preciso hacer una advertencia aquí respecto de la imposición de esas restricciones *a priori*, ". . . pues en general se desean probar las predicciones *a priori* de la teoría económica en lugar de imponerlas simplemente sobre los datos para los cuales pueden no ser válidas".[30] Sin embargo, sabemos, de la sección 8.6, cómo probar explícitamente la validez de tales restricciones.

2. **Combinación de información de corte transversal y de series de tiempo.** Una variante de la técnica de información externa o *a priori* es la combinación de datos de corte transversal y de series de tiempo, conocida como *mezcla de datos*. Suponga que deseamos estudiar la demanda de automóviles en Estados Unidos y que tenemos información de series de tiempo sobre el número de automóviles vendidos, su precio promedio y el ingreso del consumidor. Además, suponga que

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$$

donde $Y$ = número de automóviles vendidos, $P$ = precio promedio, $I$ = ingreso y $t$ = tiempo. El objetivo es estimar la elasticidad precio $\beta_2$ y la elasticidad ingreso $\beta_3$.

En la información de series de tiempo, las variables precio e ingreso tienden a ser muy colineales. Por consiguiente, si deseamos efectuar la anterior regresión, debemos enfrentar el problema usual de multicolinealidad. Tobin sugiere una salida a esto.[31] Sostiene que si hay información de corte transversal (por ejemplo, información generada a través de paneles de consumidores o estudios sindicados realizados por varias agencias privadas y estatales), puede obtenerse una estimación relativamente confiable de la elasticidad ingreso $\beta_3$, pues, con tal información, que está en un punto en el tiempo, los precios no varían mucho. Sea $\hat{\beta}_3$ la elasticidad ingreso estimada a partir de los datos de corte transversal. Con esta estimación, la anterior regresión de series de tiempo se escribe como

$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t$$

donde $Y^* = \ln Y - \hat{\beta}_3 \ln I$, es decir, $Y^*$ representa ese valor de $Y$ después de eliminarle el efecto del ingreso. Ahora se puede obtener una estimación de la elasticidad precio $\beta_2$ de la regresión anterior.

Aunque es una técnica atractiva, la mezcla de datos de series de tiempo y de corte transversal de esta forma puede crear problemas de interpretación porque se supone implícitamente que la elasticidad ingreso estimada a partir de datos de corte transversal es igual a la que se habría obtenido a partir de un análisis puro de series de tiempo.[32] Sin embargo, se ha empleado esta técnica en muchas aplicaciones y es en particular valiosa en situaciones en donde las estimaciones de corte transversal no varían sustancialmente de una sección transversal a otra. Un ejemplo de esta técnica se encuentra en el ejercicio 10.26.

3. **Eliminación de una(s) variable(s) y el sesgo de especificación.** Al enfrentar el problema de multicolinealidad grave, una de las soluciones "más simples" consiste en omitir del modelo

[30] Mark B. Stewart y Kenneth F. Wallis, *Introductory Econometrics,* 2a. ed., John Wiley & Sons, A Halstead Press Book, Nueva York, 1981, p. 154.

[31] J. Tobin, "A Statistical Demand Function for Food in the U.S.A.", *Journal of the Royal Statistical Society*, Ser. A, 1950, pp. 113-141.

[32] Hay un análisis completo y una aplicación de la técnica de datos combinados en Edwin Kuh, *Capital Stock Growth: A Micro-Econometric Approach,* North-Holland Publishing Company, Ámsterdam, 1963, capítulos 5 y 6.

una de las variables colineales. Así, en el ejemplo consumo-ingreso-riqueza, al omitir la variable riqueza, obtenemos la regresión (10.6.4), la cual muestra que mientras en el modelo original la variable ingreso no era estadísticamente significativa, ahora se vuelve "altamente" significativa.

Sin embargo, al eliminar una variable del modelo se puede incurrir en un **sesgo de especificación** o **error de especificación**. El sesgo de especificación surge de la especificación incorrecta del modelo utilizado en el análisis. Así, si la teoría económica afirma que tanto el ingreso como la riqueza deben incluirse en el modelo que explica el gasto de consumo, al eliminar la variable riqueza se incurriría en un sesgo de especificación.

Aunque estudiaremos el tema del sesgo de especificación en el capítulo 13, recuerde la idea general sobre éste dada en la sección 7.7. Si el modelo verdadero es

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

pero se ajusta de manera errónea el modelo

$$Y_i = b_1 + b_{12} X_{2i} + \hat{u}_i \tag{10.8.1}$$

se demuestra que (véase el apéndice 13A.1)

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \tag{10.8.2}$$

donde $b_{32}$ = coeficiente de la pendiente en la regresión de $X_3$ sobre $X_2$. Por consiguiente, es obvio de (10.8.2) que $b_{12}$ será una estimación sesgada de $\beta_2$ en la medida en que $b_{32}$ sea diferente de cero (se supone que $\beta_3$ es diferente de cero; en caso contrario, no tendría sentido incluir $X_3$ en el modelo original).[33] Claro está que si $b_{32}$ fuera cero, para empezar no habría problema de multicolinealidad. También es claro de (10.8.2) que si $b_{32}$ y $\beta_3$ son positivas (o ambas negativas), $E(b_{12})$ será mayor que $\beta_2$; por tanto, en promedio, $b_{12}$ sobreestimará a $\beta_2$, para ocasionar un sesgo positivo. De la misma forma, si el producto $b_{32}\beta_3$ es negativo, en promedio, $b_{12}$ subestimará a $\beta_2$, para ocasionar un sesgo negativo.

Del análisis anterior, es claro que eliminar una variable del modelo para resolver el problema de la multicolinealidad puede producir un sesgo de especificación. Por tanto, el remedio suele ser peor que la enfermedad en algunas situaciones porque, mientras que la multicolinealidad puede obstaculizar la estimación precisa de los parámetros del modelo, la omisión de una variable generaría graves equivocaciones respecto de los verdaderos valores de los parámetros. Recuerde que los estimadores de MCO son MELI a pesar de la presencia de multicolinealidad perfecta.

4. **Transformación de variables.** Suponga que tenemos información de series de tiempo sobre el gasto de consumo, el ingreso y la riqueza. Una razón de la alta multicolinealidad entre el ingreso y la riqueza en tal información es que, con el tiempo, las dos variables tienden a moverse en la misma dirección. Una forma de reducir esta dependencia es proceder de la siguiente manera.

Si la relación

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \tag{10.8.3}$$

se cumple en el periodo $t$, también debe cumplirse en el periodo $t - 1$, pues el origen del tiempo es, de todas formas, arbitrario. Por consiguiente, tenemos que:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \tag{10.8.4}$$

Si restamos (10.8.4) de (10.8.3) obtenemos

$$Y_t - Y_{t-1} = \beta_2(X_{2t} - X_{2,t-1}) + \beta_3(X_{3t} - X_{3,t-1}) + v_t \tag{10.8.5}$$

[33] Además, observe que si $b_{32}$ no se acerca a cero a medida que el tamaño de la muestra aumenta indefinidamente, entonces $b_{12}$ no sólo será sesgado, sino también inconsistente.

donde $v_t = u_t - u_{t-1}$. La ecuación (10.8.5) se conoce como la **forma en primeras diferencias** porque no se hace la regresión sobre las variables originales, sino sobre las diferencias de los valores sucesivos de dichas variables.

El modelo de regresión que utiliza primeras diferencias a menudo reduce la gravedad de la multicolinealidad porque, aunque los niveles de $X_2$ y $X_3$ estén muy correlacionados, no hay razón *a priori* para pensar que sus diferencias también lo están.

Como veremos en los capítulos que estudian la **econometría de las series de tiempo**, una ventaja incidental de la transformación de primeras diferencias consiste en que puede hacer que una serie de tiempo no estacionaria se convierta en estacionaria. En dichos capítulos veremos la importancia de las series de tiempo estacionarias. Como apreciamos en el capítulo 1, de manera muy general, una serie de tiempo, por ejemplo $Y_t$, es estacionaria si su media y varianza no cambian de manera sistemática a través del tiempo.

Otra transformación común en la práctica es la **transformación de razón**. Considere el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad \textbf{(10.8.6)}$$

donde $Y$ es el gasto de consumo en dólares reales, $X_2$ es el PIB y $X_3$ es la población total. Como el PIB y la población aumentan con el tiempo, es muy probable que estén correlacionados. Una "solución" a este problema consiste en expresar el modelo mediante una base *per cápita*; es decir, dividir (10.8.4) entre $X_3$ para obtener:

$$\frac{Y_t}{X_{3t}} = \beta_1 \left(\frac{1}{X_{3t}}\right) + \beta_2 \left(\frac{X_{2t}}{X_{3t}}\right) + \beta_3 + \left(\frac{u_t}{X_{3t}}\right) \qquad \textbf{(10.8.7)}$$

Dicha transformación tal vez reduzca la colinealidad en las variables originales.

Sin embargo, la transformación que utiliza primeras diferencias o las transformaciones de razón crean otros problemas. Por ejemplo, el término de error $v_t$ que aparece en (10.8.5) puede no satisfacer un supuesto del modelo clásico de regresión lineal, a saber, que las perturbaciones no están serialmente correlacionadas. Como veremos en el capítulo 12, si el término de perturbación $u_t$ original no está serialmente correlacionado, el término de error $v_t$ obtenido antes estará, en la mayoría de los casos, serialmente correlacionado. De nuevo, el remedio puede ser peor que la enfermedad. Además, se pierde una observación debido al procedimiento de diferenciación y, por consiguiente, los grados de libertad se reducen en 1. En una muestra pequeña esto puede ser un factor que al menos se debe considerar. Por añadidura, el procedimiento de primeras diferencias puede no ser el adecuado en los datos de corte transversal, donde no hay un ordenamiento lógico de las observaciones.

Del mismo modo, en el modelo de la razón (10.8.7), el término de error

$$\left(\frac{u_t}{X_{3t}}\right)$$

será heteroscedástico, si el término de error original $u_t$ es homoscedástico, como veremos en el capítulo 11. Una vez más, el remedio quizá resulte peor que la enfermedad de la colinealidad.

En resumen, se debe tener cuidado con las primeras diferencias o el método de la razón para transformar los datos a fin de resolver el problema de la multicolinealidad.

5. **Datos nuevos o adicionales.** Como la multicolinealidad es una característica de la muestra, es posible que en otra muestra con las mismas variables la colinealidad no sea tan grave como en la primera. A veces, con sólo aumentar el tamaño de la muestra (si esto es posible) se atenúa el problema de colinealidad. Por ejemplo, en el modelo de tres variables vimos que:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 \left(1 - r_{23}^2\right)}$$

Ahora, a medida que aumenta el tamaño de la muestra, $\sum x_{2i}^2$ por lo general aumenta. (¿Por qué?) Por consiguiente, para cualquier $r_{23}$ dado, la varianza de $\hat{\beta}_2$ disminuirá, para reducir el error estándar, lo cual permite estimar $\beta_2$ de manera más precisa.

Como ejemplo, considere la siguiente regresión del gasto de consumo $Y$ sobre el ingreso $X_2$ y la riqueza $X_3$ basada en 10 observaciones.[34]

$$\begin{aligned} \hat{Y}_i &= 24.377 \;+\; 0.8716X_{2i} - \;\;0.0349X_{3i} \\ t &= (3.875) \quad (2.7726) \quad (-1.1595) \qquad R^2 = 0.9682 \end{aligned} \tag{10.8.8}$$

El coeficiente de la riqueza en esta regresión no sólo tiene el signo equivocado, sino que estadísticamente no es significativo en el nivel de 5%. Pero cuando el tamaño de la muestra se incrementó a 40 observaciones (¿micronumerosidad?) se obtuvieron los siguientes resultados:

$$\begin{aligned} \hat{Y}_i &= \;\;2.0907 \;+\; 0.7299X_{2i} + \;0.0605X_{3i} \\ t &= (0.8713) \quad (6.0014) \quad (2.0014) \qquad R^2 = 0.9672 \end{aligned} \tag{10.8.9}$$

Ahora el coeficiente de la riqueza no sólo tiene el signo correcto, sino que es estadísticamente significativo en el nivel de 5%.

La obtención de datos adicionales o “mejores” no siempre es tan sencilla, pues, como mencionan Judge *et al.*:

> Por desgracia, muy pocas veces pueden los economistas obtener información adicional sin incurrir en altos costos, y mucho menos pueden seleccionar los valores de las variables explicativas que desean. Además, al agregar variables en situaciones no controladas, se debe tener cuidado de no agregar observaciones generadas en un proceso diferente del asociado al conjunto original de datos; es decir, se debe estar seguro de que la estructura económica asociada a las nuevas observaciones sea igual a la estructura original.[35]

6. **Reducción de la colinealidad en las regresiones polinomiales.** En la sección 7.10 estudiamos los modelos de regresión polinomial. Una característica especial de estos modelos es que la(s) variable(s) explicativa(s) aparece(n) elevada(s) a diversas potencias. Por tanto, en la función cúbica de costos totales que implica la regresión del costo total sobre la producción, la $(\text{producción})^2$ y la $(\text{producción})^3$, como en (7.10.4), los diversos términos de la producción van a estar correlacionados, lo que dificulta la estimación precisa de los diversos coeficientes de pendiente.[36] No obstante, en la práctica se ha visto que si la(s) variable(s) explicativa(s) está(n) expresada(s) en forma de desviación (es decir, desviaciones del valor medio), la multicolinealidad se reduce sustancialmente. Pero, aun entonces, el problema puede persistir,[37] en cuyo caso tal vez convenga considerar técnicas como la de los **polinomios ortogonales**.[38]

7. **Otros métodos de remediar la multicolinealidad.** Las técnicas estadísticas multivariadas como el **análisis de factores** y el de **componentes principales**, o como la **regresión en cadena**, son comunes para “resolver” el problema de la multicolinealidad. Desafortunadamente, estas técnicas están fuera del alcance de este libro, pues no pueden analizarse en forma competente sin recurrir al álgebra matricial.[39]

---

[34] El autor agradece a Albert Zucker la obtención de los resultados de las siguientes regresiones.

[35] Judge *et al., op. cit.,* p. 625. Véase también la sección 10.9.

[36] Como ya mencionamos, puesto que la relación entre $X$, $X^2$ y $X^3$ es no lineal, las regresiones polinomiales no violan el supuesto de no multicolinealidad del modelo clásico, en estricto sentido.

[37] Véase R.A. Bradley y S.S. Srivastava, “Correlation and Polynomial Regression”, *American Statistician,* vol. 33, 1979, pp. 11-14.

[38] Véase Norman Draper y Harry Smith, *Applied Regression Analysis,* 2a. ed., John Wiley & Sons, Nueva York, 1981, pp. 266-274.

[39] Una explicación sencilla de estas técnicas, desde un punto de vista aplicado, se encuentra en Samprit Chatterjee y Bertram Price, *Regression Analysis by Example,* John Wiley & Sons, Nueva York, 1977, capítulos 7 y 8. Véase también H.D. Vinod, “A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares”, *Review of Economics and Statistics,* vol. 60, febrero de 1978, pp. 121-131.

## 10.9 ¿Es la multicolinealidad necesariamente mala? Quizá no, si el objetivo es sólo la predicción

Dijimos que si el único propósito del análisis de regresión es el pronóstico o la predicción, la multicolinealidad no es un problema grave, pues, entre más alta sea la $R^2$, mejor será la predicción.[40] Pero esto sucede "… siempre que los valores de las variables explicativas, para los cuales se desean las predicciones, obedezcan las mismas dependencias lineales casi exactas de la matriz $X$ [de datos] del diseño original".[41] Por tanto, si en una regresión estimada se encuentra que $X_2 = 2X_3$ aproximadamente, entonces, en una muestra futura para pronosticar $Y$, $X_2$ también debe ser aproximadamente igual a $2X_3$, condición difícil de cumplir en la práctica (véase la nota 35), en cuyo caso la predicción será cada vez más incierta.[42] Más aún, si el objetivo del análisis no es sólo la predicción sino también la estimación confiable de los parámetros, la presencia de una alta multicolinealidad puede ser un problema porque, como vimos, genera grandes errores estándar en los estimadores.

Sin embargo, existen situaciones en las cuales la multicolinealidad puede no representar un problema grave. Es el caso en el cual se tiene una $R^2$ elevada y los coeficientes de regresión son significativos individualmente como lo demuestran los altos valores $t$. Aun así, los diagnósticos de multicolinealidad, por ejemplo el índice de condición, indican que los datos presentan colinealidad grave. ¿Cuándo puede presentarse tal situación? Como menciona Johnston:

> Esto sucede si los coeficientes individuales resultan estar numéricamente muy por encima del valor verdadero, de forma que el efecto siga visible, a pesar de los errores estándar inflados y/o debido a que el valor verdadero es en sí mismo tan grande que, aunque se obtenga una estimación subestimada, continúe siendo significativa.[43]

## 10.10 Ejemplo ampliado: los datos Longley

Concluimos este capítulo con el análisis de los datos recopilados por Longley.[44] Aunque se obtuvieron originalmente para evaluar la exactitud del cálculo computacional de las estimaciones de mínimos cuadrados de varios paquetes de software, los datos Longley se convirtieron en ejemplo para ilustrar diversos problemas econométricos, como la multicolinealidad. Los datos se reproducen en la tabla 10.8, y son series de tiempo de 1947 a 1962, donde $Y$ = número de personas con trabajo (en miles), $X_1$ = índice implícito de deflación de precios para el PIB, $X_2$ = PIB (en millones de dólares), $X_3$ = número de desempleados (en miles), $X_4$ = número de personas enlistadas en las fuerzas armadas, $X_5$ = población no institucionalizada mayor de 14 años de edad y $X_6$ = año (igual a 1 para 1947, 2 para 1948 y 16 para 1962).

---

[40] Véase R.C. Geary, "Some Results about Relations Between Stochastic Variables: A Discussion Document", *Review of International Statistical Institute*, vol. 31, 1963, pp. 163-181.

[41] Judge *et al., op. cit.,* p. 619. También encontrará en esta página la prueba de que, a pesar de la colinealidad, se obtienen mejores predicciones medias si la estructura de colinealidad existente también persiste en las muestras futuras.

[42] Para un excelente análisis véase E. Malinvaud, *Statistical Methods of Econometrics*, 2a. ed., North-Holland Publishing Company, Ámsterdam, 1970, pp. 220-221.

[43] J. Johnston, *Econometric Methods*, 3a. ed., McGraw-Hill, Nueva York, 1984, p. 249.

[44] J. Longley, "An Appraisal of Least-Squares Programs from the Point of User", *Journal of the American Statistical Association*, vol. 62, 1967, pp. 819-841.

**TABLA 10.8**
**Datos Longley**

Fuente: J. Longley, "An Appraisal of Least-Squares Programs from the Point of the User", *Journal of the American Statistical Association*, vol. 62, 1967, pp. 819-841.

| Observación | Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Tiempo |
|---|---|---|---|---|---|---|---|
| 1947 | 60 323 | 830 | 234 289 | 2 356 | 1 590 | 107 608 | 1 |
| 1948 | 61 122 | 885 | 259 426 | 2 325 | 1 456 | 108 632 | 2 |
| 1949 | 60 171 | 882 | 258 054 | 3 682 | 1 616 | 109 773 | 3 |
| 1950 | 61 187 | 895 | 284 599 | 3 351 | 1 650 | 110 929 | 4 |
| 1951 | 63 221 | 962 | 328 975 | 2 099 | 3 099 | 112 075 | 5 |
| 1952 | 63 639 | 981 | 346 999 | 1 932 | 3 594 | 113 270 | 6 |
| 1953 | 64 989 | 990 | 365 385 | 1 870 | 3 547 | 115 094 | 7 |
| 1954 | 63 761 | 1 000 | 363 112 | 3 578 | 3 350 | 116 219 | 8 |
| 1955 | 66 019 | 1 012 | 397 469 | 2 904 | 3 048 | 117 388 | 9 |
| 1956 | 67 857 | 1 046 | 419 180 | 2 822 | 2 857 | 118 734 | 10 |
| 1957 | 68 169 | 1 084 | 442 769 | 2 936 | 2 798 | 120 445 | 11 |
| 1958 | 66 513 | 1 108 | 444 546 | 4 681 | 2 637 | 121 950 | 12 |
| 1959 | 68 655 | 1 126 | 482 704 | 3 813 | 2 552 | 123 366 | 13 |
| 1960 | 69 564 | 1 142 | 502 601 | 3 931 | 2 514 | 125 368 | 14 |
| 1961 | 69 331 | 1 157 | 518 173 | 4 806 | 2 572 | 127 852 | 15 |
| 1962 | 70 551 | 1 169 | 554 894 | 4 007 | 2 827 | 130 081 | 16 |

Suponga que nuestro objetivo es predecir $Y$ con base en las seis variables $X$. Mediante el software *EViews6* obtenemos los siguientes resultados de la regresión:

Variable dependiente: $Y$
Muestra: 1947-1962

| Variable | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| $C$ | -3482259. | 890420.4 | -3.910803 | 0.0036 |
| $X_1$ | 15.06187 | 84.91493 | 0.177376 | 0.8631 |
| $X_2$ | -0.035819 | 0.033491 | -1.069516 | 0.3127 |
| $X_3$ | -2.020230 | 0.488400 | -4.136427 | 0.0025 |
| $X_4$ | -1.033227 | 0.214274 | -4.821985 | 0.0009 |
| $X_5$ | -0.051104 | 0.226073 | -0.226051 | 0.8262 |
| $X_6$ | 1829.151 | 455.4785 | 4.015890 | 0.0030 |

| | | | |
|---|---|---|---|
| $R$ cuadrada | 0.995479 | Media de la variable dependiente | 65317.00 |
| $R$ cuadrada ajustada | 0.992465 | Desviación estándar de la variable dependiente | 3511.968 |
| Error estándar de la regresión | 304.8541 | Criterio de información de Akaike | 14.57718 |
| Suma de cuadrados residual | 836424.1 | Criterio de Schwarz | 14.91519 |
| Log verosimilitud | -109.6174 | Estadístico $F$ | 330.2853 |
| Estadístico de Durbin-Watson | 2.559488 | Probabilidad (estadístico $F$) | 0.000000 |

A primera vista, dichos resultados sugieren que se tiene un problema de colinealidad, pues el valor $R^2$ es muy alto; sin embargo, unas cuantas variables son estadísticamente no significativas ($X_1$, $X_2$ y $X_5$), lo cual constituye un síntoma característico de multicolinealidad. Para arrojar más luz a este problema, en la tabla 10.9 se presentan las intercorrelaciones entre las seis regresoras.

Esta tabla suministra lo que se llama **matriz de correlación**. En la tabla, las entradas de la diagonal principal (las que van desde la esquina superior izquierda hacia la esquina inferior derecha) suministran la correlación de una variable consigo misma, la cual por definición siempre es 1; además, las entradas fuera de la diagonal principal son las parejas de correlaciones entre las variables $X$. El primer renglón de esta tabla proporciona la correlación de $X_1$ con las otras varia-

**TABLA 10.9**
**Intercorrelaciones**

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| $X_1$ | 1.000000 | 0.991589 | 0.620633 | 0.464744 | 0.979163 | 0.991149 |
| $X_2$ | 0.991589 | 1.000000 | 0.604261 | 0.446437 | 0.991090 | 0.995273 |
| $X_3$ | 0.620633 | 0.604261 | 1.000000 | −0.177421 | 0.686552 | 0.668257 |
| $X_4$ | 0.464744 | 0.446437 | −0.177421 | 1.000000 | 0.364416 | 0.417245 |
| $X_5$ | 0.979163 | 0.991090 | 0.686552 | 0.364416 | 1.000000 | 0.993953 |
| $X_6$ | 0.991149 | 0.995273 | 0.668257 | 0.417245 | 0.993953 | 1.000000 |

bles $X$. Por ejemplo, 0.991589 es la correlación entre $X_1$ y $X_2$; 0.620633 es la correlación entre $X_1$ y $X_3$, y así sucesivamente.

Como se ve, varias de estas correlaciones a pares son muy altas, lo cual sugiere que quizá haya un grave problema de colinealidad. Por supuesto, debe recordarse la advertencia anterior de que tales correlaciones a pares tal vez sean una condición suficiente, pero no necesaria, para la multicolinealidad.

Con objeto de aclarar más la naturaleza del problema de la multicolinealidad, observe las regresiones auxiliares; es decir, la regresión de cada variable $X$ sobre las restantes variables $X$. Para ahorrar espacio, se presentarán sólo los valores $R^2$ obtenidos con base en esas regresiones, las cuales se listan en la tabla 10.10. Como los valores $R^2$ de las regresiones auxiliares son muy altos (con la posible excepción de la regresión de $X_4$) sobre las restantes variables $X$, al parecer existe un grave problema de colinealidad. La misma información se obtiene a partir de los factores de tolerancia. Como ya mencionamos, mientras más cercano a cero esté el factor de tolerancia, mayor será la evidencia de colinealidad.

Al aplicar la regla práctica de Klein observamos que los valores $R^2$ obtenidos de las regresiones auxiliares exceden el valor general $R^2$ (es decir, el que se obtuvo de la regresión de $Y$ sobre todas las variables $X$), que es igual a 0.9954, en 3 de 6 regresiones auxiliares, lo cual de nuevo sugiere que sin duda los datos Longley están plagados del problema de multicolinealidad. A propósito, si aplica la prueba $F$ dada en (10.7.3), el lector debe verificar que todos los valores $R^2$ dados en las tablas anteriores son estadística y significativamente diferentes de cero.

Ya observamos que los estimadores de MCO y sus errores estándar son sensibles a los pequeños cambios en los datos. En el ejercicio 10.32 se pide al lector que vuelva a efectuar la regresión de $Y$ sobre cada una de las seis variables $X$, pero que elimine las últimas observaciones; es decir, que haga la regresión para el periodo 1947-1961. Verá cómo cambian los resultados de la regresión al eliminar las observaciones de un solo año.

Ahora que establecimos que existe un problema de multicolinealidad, ¿qué acciones correctivas pueden llevarse a cabo? Reconsidere el modelo original. En primer lugar, el PIB puede expresarse no en términos nominales, sino en términos reales, lo cual se realiza al dividir el PIB nominal entre el índice de deflación del precio implícito. En segundo lugar, en vista de que la población no institucional mayor de 14 años aumenta con el tiempo debido al crecimiento natural de la población, estará muy correlacionada con el tiempo, la variable $X_6$ del modelo. Por tanto, en lugar de conservar esas dos variables, mantenemos la variable $X_5$ y desechamos $X_6$. En tercer

**TABLA 10.10**
**Valores $R^2$ obtenidos de regresiones auxiliares**

| Variable dependiente | Valor de $R^2$ | Tolerancia (TOL) = 1 − $R^2$ |
|---|---|---|
| $X_1$ | 0.9926 | 0.0074 |
| $X_2$ | 0.9994 | 0.0006 |
| $X_3$ | 0.9702 | 0.0298 |
| $X_4$ | 0.7213 | 0.2787 |
| $X_5$ | 0.9970 | 0.0030 |
| $X_6$ | 0.9986 | 0.0014 |

lugar, no hay ninguna razón de peso para incluir $X_3$, el número de personas desempleadas; quizá la tasa de desempleo fuese una mejor medida de las condiciones del mercado de trabajo; sin embargo, no hay ningún dato al respecto. Por consiguiente, eliminamos la variable $X_3$. Con estos cambios obtenemos los siguientes resultados de la regresión (PIBR = PIB real):[45]

Variable dependiente: *Y*
Muestra: 1947-1962

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C* | 65720.37 | 10624.81 | 6.185558 | 0.0000 |
| PIBR | 9.736496 | 1.791552 | 5.434671 | 0.0002 |
| $X_4$ | -0.687966 | 0.322238 | -2.134965 | 0.0541 |
| $X_5$ | -0.299537 | 0.141761 | -2.112965 | 0.0562 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.981404 | Media de la variable dependiente | 65317.00 |
| *R* cuadrada ajustada | 0.976755 | Desviación estándar de la variable dependiente | 3511.968 |
| Error estándar de la regresión | 534.4492 | Criterio de información de Akaike | 15.61641 |
| Suma de cuadrados residual | 3440470. | Criterio de Schwarz | 15.80955 |
| Log verosimilitud | -120.9313 | Estadístico *F* | 211.0972 |
| Estadístico de Durbin-Watson | 1.654069 | Probabilidad (estadístico *F*) | 0.000000 |

Aunque $R^2$ disminuyó un poco en comparación con la $R^2$ original, aún es muy alta. Ahora todos los coeficientes estimados son significativos y sus signos tienen sentido desde el punto de vista económico.

Dejamos al lector encontrar otros modelos y observar la forma en que cambian los resultados. También tenga en cuenta la advertencia anterior respecto de la utilización del método de la razón para transformar los datos con el objeto de resolver el problema de la colinealidad. Volveremos a esta cuestión en el capítulo 11.

## Resumen y conclusiones

1. Un supuesto del modelo clásico de regresión lineal es que no haya multicolinealidad entre las variables explicativas, las *X*. Interpretada en términos generales, la multicolinealidad se refiere a una situación en la cual existe una relación lineal exacta o aproximadamente exacta entre las variables *X*.
2. Las consecuencias de la multicolinealidad son las siguientes: si existe colinealidad perfecta entre las *X*, sus coeficientes de regresión son indeterminados y sus errores estándar no están definidos. Si la colinealidad es alta pero no perfecta, es posible la estimación de los coeficientes de regresión, pero sus errores estándar tienden a ser grandes. Como resultado, los valores poblacionales de los coeficientes no pueden estimarse en forma precisa; sin embargo, si el objetivo es estimar combinaciones lineales de estos coeficientes, las *funciones estimables*, esto se logra aun en presencia de multicolinealidad perfecta.
3. Aunque no hay métodos seguros para detectar la colinealidad, existen diversos indicadores, como los siguientes:
   *a*) El signo más claro de multicolinealidad es cuando $R^2$ es muy alta pero ninguno de los coeficientes de regresión es estadísticamente significativo con base en la prueba *t* convencional. Este caso es, por supuesto, extremo.

[45] El coeficiente de correlación entre $X_5$ y $X_6$ es de casi 0.9939, una correlación muy alta en verdad.

*b*) En los modelos con apenas dos variables explicativas, puede tenerse una idea relativamente buena de la colinealidad mediante el examen del coeficiente de correlación de orden cero, o simple, entre las dos variables. Si esta correlación es alta, la multicolinealidad suele ser la culpable.

*c*) Sin embargo, los coeficientes de correlación de orden cero pueden ser malos indicadores en modelos con más de dos variables $X$, pues es posible tener correlaciones bajas de orden cero y encontrar aún alta multicolinealidad. En estas situaciones puede ser necesario examinar los coeficientes de correlación parcial.

*d*) Si $R^2$ es alta pero las correlaciones parciales son bajas, la multicolinealidad es una posibilidad. Aquí hay una o más variables que pueden ser superfluas. Pero si $R^2$ es alta y las correlaciones parciales también son altas, la multicolinealidad puede no ser fácilmente detectable. También, como señalan C. Robert Wichers, Krishna Kumar, John O'Hagan y Brendan McCabe, hay algunos problemas estadísticos con la prueba de correlación parcial sugerida por Farrar y Glauber.

*e*) Por consiguiente, se puede hacer la regresión de cada variable $X_i$ sobre las variables $X$ restantes en el modelo y encontrar los coeficientes de determinación correspondientes $R_i^2$. Una $R_i^2$ elevada sugeriría que $X_i$ está muy correlacionado con el resto de las $X$. Así, se puede eliminar esa $X_i$ del modelo siempre y cuando no conduzca a un sesgo de especificación grave.

4. La detección de multicolinealidad es la mitad de la batalla. La otra mitad es hallar la forma de deshacerse del problema. Nuevamente, no existen métodos seguros, sólo unas cuantas reglas prácticas, algunas de las cuales son las siguientes: 1) utilizar información obtenida *a priori* o externa al modelo, 2) combinar información de corte transversal y de series de tiempo, 3) omitir una variable si es muy colineal, 4) transformar los datos y 5) obtener datos adicionales o nuevos. Naturalmente, saber qué regla funciona en la práctica depende de la naturaleza de la información y de la gravedad del problema de colinealidad.
5. Mencionamos el papel de la multicolinealidad en la predicción y señalamos que, a menos que la estructura colineal continúe en muestras futuras, es peligroso utilizar para fines de proyección una regresión estimada que haya sido contaminada por la multicolinealidad.
6. Aunque la multicolinealidad ha recibido extensa (algunos dirían excesiva) atención en la teoría, un problema igualmente importante en la investigación empírica es el de la micronumerosidad, o pequeñez del tamaño de la muestra. De acuerdo con Goldberger: "Cuando un artículo de investigación acusa la presencia de multicolinealidad, los lectores deben ver si esa queja sería convincente si se sustituyera el concepto de 'micronumerosidad' por el de 'multicolinealidad' ".[46] Él sugiere que el lector es quien debe decidir cuán pequeña puede ser $n$, el número de observaciones, antes de concluir que se tiene un problema de muestra pequeña, de la misma forma que decide cuán alto es un valor de $R^2$ en una regresión auxiliar antes de declarar que el problema de colinealidad es muy grave.

# EJERCICIOS

## Preguntas

10.1. En el modelo de regresión lineal de $k$ variables, hay $k$ ecuaciones normales para estimar las $k$ incógnitas. Estas ecuaciones normales están dadas en el **apéndice C**. Suponga que $X_k$ es una combinación lineal perfecta de las variables $X$ restantes. ¿Cómo se demostraría que en este caso es imposible estimar los $k$ coeficientes de regresión?

[46] Goldberger, *op. cit.*, p. 250.

**TABLA 10.11**

| *Y* | $X_2$ | $X_3$ |
|---|---|---|
| −10 | 1 | 1 |
| −8 | 2 | 3 |
| −6 | 3 | 5 |
| −4 | 4 | 7 |
| −2 | 5 | 9 |
| 0 | 6 | 11 |
| 2 | 7 | 13 |
| 4 | 8 | 15 |
| 6 | 9 | 17 |
| 8 | 10 | 19 |
| 10 | 11 | 21 |

10.2. Considere el conjunto de datos hipotéticos de la tabla 10.11. Suponga que desea ajustar el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

a los datos.

*a*) ¿Puede estimar las tres incógnitas? ¿Por qué?

*b*) Si no se puede hacer, ¿qué funciones lineales de estos parámetros, las funciones estimables, puede estimar? Muestre los cálculos necesarios.

10.3. Consulte el ejemplo de la mortalidad infantil analizado en el capítulo 8 (ejemplo 8.1). Dicho ejemplo implicó hacer la regresión de la tasa de mortalidad infantil (MI) sobre el PIB *per cápita* (PIBPC) y la tasa de alfabetización de las mujeres (TAM). Ahora, suponga que añadimos la variable tasa de fecundidad total (TFT). Lo anterior da los siguientes resultados de la regresión:

Variable dependiente: MI

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C* | 168.3067 | 32.89165 | 5.117003 | 0.0000 |
| PIBPC | -0.005511 | 0.001878 | -2.934275 | 0.0477 |
| TAM | -1.768029 | 0.248017 | -7.128663 | 0.0000 |
| TFT | 12.86864 | 4.190533 | 3.070883 | 0.0032 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.747372 | Media de la variable dependiente | 141.5000 |
| *R* cuadrada ajustada | 0.734740 | Desviación estándar de la variable dependiente | 75.97807 |
| Error estándar de la regresión | 39.13127 | Criterio de información de Akaike | 10.23218 |
| Suma de cuadrados residual | 91875.38 | Criterio de Schwarz | 10.36711 |
| Log verosimilitud | -323.4298 | Estadístico *F* | 59.16767 |
| Estadístico de Durbin-Watson | 2.170318 | Probabilidad (estadístico *F*) | 0.000000 |

*a*) Compare estos resultados de la regresión con los obtenidos en la ecuación (8.1.4). ¿Qué cambios observa? ¿Cómo los explica?

*b*) ¿Vale la pena añadir la variable TFT al modelo? ¿Por qué?

*c*) Como todos los coeficientes *t* individuales son estadísticamente significativos, ¿podemos decir que no existe un problema de colinealidad en el presente caso?

10.4. Si la relación $\lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i} = 0$ se mantiene para todos los valores de $\lambda_1$, $\lambda_2$ y $\lambda_3$, estime $r_{12.3}$, $r_{13.2}$ y $r_{23.1}$. Encuentre también $R^2_{1.23}$, $R^2_{2.13}$ y $R^2_{3.12}$. ¿Cuál es el grado de

multicolinealidad en esta situación? *Nota:* $R^2_{1.2\,3}$ es el coeficiente de determinación en la regresión de $Y$ sobre $X_2$ y $X_3$. Otros valores $R^2$ deben interpretarse en forma similar.

10.5. Considere el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \beta_5 X_{t-3} + \beta_6 X_{t-4} + u_t$$

donde $Y =$ consumo, $X =$ ingreso y $t =$ tiempo. El modelo anterior postula que el gasto de consumo en el tiempo $t$ es función no sólo del ingreso en el tiempo $t$, sino también del ingreso en periodos anteriores. Por tanto, el gasto de consumo en el primer trimestre de 2000 es función del ingreso en ese trimestre y en los cuatro trimestres de 1999. Tales modelos se denominan modelos de **rezago distribuido** y los veremos en un capítulo posterior.

*a*) ¿Esperaría multicolinealidad en tales modelos y por qué?

*b*) Si espera colinealidad, ¿cómo resolvería el problema?

10.6. Considere el ejemplo ilustrativo de la sección 10.6 (ejemplo 10.1). ¿Cómo interpretaría la diferencia en la propensión marginal a consumir obtenida de (10.6.1) y (10.6.4)?

10.7. En los datos que comprenden series de tiempo económicas, como PNB, oferta monetaria, precios, ingreso, desempleo, etc., suele sospecharse la presencia de multicolinealidad. ¿Por qué?

10.8. Suponga en el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

que $r_{23}$, el coeficiente de correlación entre $X_2$ y $X_3$, es cero. Por consiguiente, le sugieren que efectúe las siguientes regresiones:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i}$$
$$Y_i = \gamma_1 + \gamma_3 X_{3i} + u_{2i}$$

*a*) ¿Será $\hat{\alpha}_2 = \hat{\beta}_2$ y $\hat{\gamma}_3 = \hat{\beta}_3$? ¿Por qué?

*b*) ¿Será $\hat{\beta}_1$ igual $\hat{\alpha}_1$ o $\hat{\gamma}_1$, o a alguna combinación de éstos?

*c*) ¿Será $\text{var}(\hat{\beta}_2) = \text{var}(\hat{\alpha}_2)$ y $\text{var}(\hat{\beta}_3) = \text{var}(\hat{\gamma}_3)$?

10.9. Consulte el ejemplo ilustrativo del capítulo 7, en el cual ajustamos la función de producción Cobb-Douglas al sector manufacturero de los 50 estados y el Distrito de Columbia de Estados Unidos para 2005. Los resultados de la regresión dados en (7.9.4) muestran que los coeficientes del trabajo y del capital son estadísticamente significativos en lo individual.

*a*) Examine si las variables trabajo y capital están muy correlacionadas.

*b*) Si la respuesta a *a*) es afirmativa, ¿eliminaría, por ejemplo, la variable trabajo del modelo y efectuaría la regresión de la variable producción sobre el insumo capital solamente?

*c*) Si hace lo anterior, ¿en qué clase de sesgo de especificación se incurre? Descubra la naturaleza de este sesgo.

10.10. Consulte el ejemplo 7.4. Para este problema, la matriz de correlación es la siguiente:

| | $X_i$ | $X_i^2$ | $X_i^3$ |
|---|---|---|---|
| $X_i$ | 1 | 0.9742 | 0.9284 |
| $X_i^2$ | | 1.0 | 0.9872 |
| $X_i^3$ | | | 1.0 |

*a*) "Como las correlaciones de orden cero son muy elevadas debe existir multicolinealidad grave." Comente.

*b*) ¿Eliminaría del modelo las variables $X_i^2$ y $X_i^3$?

*c*) Si elimina las variables anteriores, ¿que sucederá con el valor del coeficiente de $X_i$?

10.11. *Regresión por pasos*. Al decidir sobre el "mejor" conjunto de variables explicativas para un modelo de regresión, los investigadores a menudo siguen el método de regresión por pasos. En este método se introducen, una por una, las variables $X$ (**regresión por pasos hacia delante**) o se incluyen todas las variables $X$ posibles en una regresión múltiple y se rechazan una a la vez (**regresión por pasos hacia atrás**). La decisión de aumentar o eliminar una variable suele tomarse con base en la contribución de esa variable a la SCE, a juicio de la prueba $F$. Con todo lo que sabe sobre multicolinealidad, ¿recomendaría alguno de estos procedimientos? ¿Por qué?*

10.12. Establezca si las siguientes afirmaciones son verdaderas, falsas o inciertas, y *justifique* la respuesta.

*a*) A pesar de la multicolinealidad perfecta, los estimadores de MCO son MELI.

*b*) En los casos de alta multicolinealidad, no es posible evaluar la significancia individual de uno o más coeficientes de regresión parcial.

*c*) Si una regresión auxiliar muestra que una $R_i^2$ particular es alta, hay evidencia clara de alta colinealidad.

*d*) Las correlaciones altas entre parejas de regresoras no sugieren una alta multicolinealidad.

*e*) La multicolinealidad es inofensiva si el objetivo del análisis es sólo la predicción.

*f*) Entre mayor sea el FIV, *ceteris paribus*, más grandes serán las varianzas de los estimadores de MCO.

*g*) La tolerancia (TOL) es una medida de multicolinealidad mejor que el FIV.

*h*) No podrá obtener un valor $R^2$ elevado en una regresión múltiple si todos los coeficientes parciales de pendiente no son estadísticamente significativos, en lo *individual*, con base en la prueba $t$ usual.

*i*) En la regresión de $Y$ sobre $X_2$ y $X_3$, suponga que hay poca variabilidad en los valores de $X_3$. Esto aumentaría var $(\hat{\beta}_3)$. En el extremo, si todas las $X_3$ fueran idénticas, var $(\hat{\beta}_3)$ sería infinita.

10.13. *a*) Demuestre que si $r_{1i} = 0$ para $i = 2, 3, \ldots, k$ entonces

$$R_{1.23\ldots k} = 0$$

*b*) ¿Qué importancia tiene este hallazgo para la regresión de la variable $X_1$ $(= Y)$ sobre $X_2, X_3, \ldots, X_k$?

10.14. Suponga que todos los coeficientes de correlación de orden cero de $X_1$ $(= Y), X_2, \ldots, X_k$ son iguales a $r$.

*a*) ¿Cuál es el valor de $R^2_{1.23\ldots k}$?

*b*) ¿Cuáles son los valores de los coeficientes de correlación de primer orden?

**10.15. En notación matricial se demuestra (**apéndice C**) que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

*a*) ¿Qué sucede con $\hat{\boldsymbol{\beta}}$ cuando hay colinealidad perfecta entre las $X$?

*b*) ¿Cómo sabe si existe colinealidad perfecta?

* Compruebe si su razonamiento está de acuerdo con el de Arthur S. Goldberger y D.B. Jochems, "Note on Stepwise Least-Squares", *Journal of the American Statistical Association*, vol. 56, marzo de 1961, pp. 105-110.

** Opcional.

*10.16. Con notación matricial se demuestra que

$$\text{var–cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

¿Qué le sucede a esta matriz var-cov

*a*) cuando hay multicolinealidad perfecta?

*b*) cuando la colinealidad es alta pero no perfecta?

*10.17. Considere la siguiente **matriz de correlación**:

$$\mathbf{R} = \begin{array}{c} \\ X_2 \\ X_3 \\ \\ X_k \end{array} \begin{array}{c} \begin{array}{cccc} X_2 & X_3 & \cdots & X_k \end{array} \\ \begin{bmatrix} 1 & r_{23} & \cdots & r_{2k} \\ r_{32} & 1 & \cdots & r_{3k} \\ \cdots & \cdots & \cdots & \\ r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix} \end{array}$$

Describa cómo calcularía, con base en la matriz de correlación, si *a*) hay colinealidad perfecta, *b*) hay colinealidad menos que perfecta y *c*) las $X$ no están correlacionadas.

*Sugerencia:* Puede utilizar $|\mathbf{R}|$ para responder estas preguntas, donde $|\mathbf{R}|$ denota el determinante de $\mathbf{R}$.

*10.18. *Variables explicativas ortogonales*. Suponga que en el siguiente modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

$X_2$ a $X_k$ no están correlacionadas. Tales variables se denominan **variables ortogonales**. Si éste es el caso:

*a*) ¿Cuál será la estructura de la matriz $(\mathbf{X}'\mathbf{X})$?

*b*) ¿Cómo obtiene $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$?

*c*) ¿Cuál será la naturaleza de la matriz var-cov de $\hat{\boldsymbol{\beta}}$?

*d*) Suponga que efectuó la regresión y luego desea introducir otra variable ortogonal, por ejemplo, $X_{k+1}$, en el modelo. ¿Debe recalcular todos los coeficientes anteriores $\hat{\beta}_1$ a $\hat{\beta}_k$? ¿Por qué?

10.19. Considere el siguiente modelo:

$$\text{PNB}_t = \beta_1 + \beta_2 \text{M}_t + \beta_3 \text{M}_{t-1} + \beta_4(\text{M}_t - \text{M}_{t-1}) + u_t$$

donde $\text{PNB}_t$ = PNB en el periodo $t$, $\text{M}_t$ = oferta monetaria en el periodo $t$, $\text{M}_{t-1}$ = oferta monetaria en el periodo $(t-1)$ y $(\text{M}_t - \text{M}_{t-1})$ = cambio en la oferta monetaria entre el tiempo $t$ y el tiempo $(t-1)$. Este modelo postula entonces que el nivel del PNB en el periodo $t$ es una función de la oferta monetaria en el tiempo $t$ y en el tiempo $(t-1)$, lo mismo que del cambio en la oferta monetaria entre estos periodos.

*a*) Si supone que tiene la información para estimar el modelo anterior, ¿podría estimar todos los coeficientes de este modelo? ¿Por qué?

*b*) De no ser posible, ¿qué coeficientes pueden estimarse?

*c*) Suponga que el término $\beta_3\text{M}_{t-1}$ estuviera ausente del modelo. ¿Su respuesta a *a*) sería la misma?

*d*) Repita *c*) pero ahora suponga que el término $\beta_2\text{M}_t$ está ausente del modelo.

* Opcional.

10.20. Muestre que (7.4.7) y (7.4.8) también se expresan como

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right)\left(1 - r_{23}^2\right)}$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right)\left(1 - r_{23}^2\right)}$$

donde $r_{23}$ es el coeficiente de correlación entre $X_2$ y $X_3$.

10.21. Con (7.4.12) y (7.4.15) muestre que cuando hay colinealidad perfecta, las varianzas de $\hat{\beta}_2$ y $\hat{\beta}_3$ son infinitas.

10.22. Verifique que los errores estándar de las sumas de los coeficientes de pendiente, estimados a partir de (10.5.6) y (10.5.7), son 0.1549 y 0.1825, respectivamente (consulte la sección 10.5).

10.23. Para el modelo de regresión con $k$ variables puede mostrarse que la varianza del $k$-ésimo coeficiente de regresión parcial ($k = 2, 3, \ldots, K$) dado en la ecuación (7.5.6) también se expresa como*

$$\operatorname{var}(\hat{\beta}_k) = \frac{1}{n-k}\frac{\sigma_y^2}{\sigma_k^2}\left(\frac{1-R^2}{1-R_k^2}\right)$$

donde $\sigma_y^2$ = varianza de $Y$, $\sigma_k^2$ = varianza de la $k$-ésima variable explicativa, $R_k^2 = R^2$ de la regresión de $X_k$ sobre las variables $X$ restantes y $R^2$ = coeficiente de determinación de la regresión múltiple, es decir, la regresión de $Y$ sobre todas las variables $X$.

*a*) ¿Qué sucede con $\operatorname{var}(\hat{\beta}_k)$ si aumenta $\sigma_k^2$, y todo lo demás se mantiene constante? ¿Cuáles son las implicaciones para el problema de multicolinealidad?

*b*) ¿Qué sucede con la fórmula anterior cuando la colinealidad es perfecta?

*c*) Responda si es cierto o es falso: "La varianza de $\hat{\beta}_k$ disminuye a medida que $R^2$ aumenta, de manera que el efecto de una $R_k^2$ alta puede compensarse por una $R^2$ alta".

10.24. Con base en la información anual para el sector manufacturero de Estados Unidos de 1899 a 1922, Dougherty obtuvo los siguientes resultados de regresión:†

$$\begin{aligned}\widehat{\log Y} &= 2.81 - 0.53 \log K + 0.91 \log L + 0.047t \\ \text{ee} &= (1.38) \quad (0.34) \qquad (0.14) \qquad (0.021) \\ & \qquad R^2 = 0.97 \qquad F = 189.8\end{aligned} \tag{1}$$

donde $Y$ = índice de producción real, $K$ = índice de insumo capital real, $L$ = índice de insumo trabajo real, $t$ = tiempo o tendencia.

Con la misma información, obtuvo también la siguiente regresión:

$$\begin{aligned}\widehat{\log (Y) L} &= -0.11 + 0.11 \log (K/L) + 0.006t \\ \text{ee} &= (0.03) \quad (0.15) \qquad (0.006) \\ & \qquad R^2 = 0.65 \qquad F = 19.5\end{aligned} \tag{2}$$

* Esta fórmula proviene de R. Stone, "The Analysis of Market Demand", *Journal of the Royal Statistical Society*, vol. B7, 1945, p. 297. Recuerde también (7.5.6). Para un análisis más detallado, véase Peter Kennedy, *A Guide to Econometrics*, 2a. ed., The MIT Press, Cambridge, Massachusetts, 1985, p. 156.

† Christopher Dougherty, *Introduction to Econometrics*, Oxford University Press, Nueva York, 1992, pp. 159-160.

*a*) ¿Existe multicolinealidad en la regresión (1)? ¿Cómo sabe?

*b*) En la regresión (1), ¿cuál es el signo *a priori* de log *K*? ¿Concuerdan los resultados con esta expectativa? ¿Por qué?

*c*) ¿Cómo justificaría la forma funcional de la regresión (1)? (*Sugerencia:* Función de producción de Cobb-Douglas.)

*d*) Interprete la regresión (1). ¿Cuál es el papel de la variable de tendencia en esta regresión?

*e*) ¿Cuál es la lógica de la estimación de la regresión (2)?

*f*) Si hubo multicolinealidad en la regresión (1), ¿se vio reducida por la regresión (2)? ¿Cómo sabe?

*g*) Si la regresión (2) es una versión restringida de la regresión (1), ¿qué restricción impone el autor? (*Sugerencia:* Rendimientos a escala.) ¿Cómo sabe si esta restricción es válida? ¿Qué prueba utiliza? Muestre todos sus cálculos.

*h*) ¿Son comparables los valores $R^2$ de las dos regresiones? ¿Por qué? ¿Cómo pueden hacerse comparables, si no lo son en la forma actual?

10.25. Evalúe de manera crítica los siguientes enunciados:

*a*) "De hecho, la multicolinealidad no es un error del modelo. Es una condición de la insuficiencia de datos."*

*b*) "Si no es factible obtener más datos, debe aceptarse que los datos con que se cuenta tienen una cantidad limitada de información, y, en consecuencia, se debe simplificar el modelo. Tratar de estimar modelos que a su vez son muy complicados es uno de los errores más comunes que cometen los econometristas aplicados inexpertos."**

*c*) "Para los investigadores resulta común afirmar que existe multicolinealidad siempre que los signos que se supusieron no se encuentran en los resultados de la regresión, cuando las variables que se sabe *a priori* que son importantes tienen valores *t* insignificantes, o cuando diversos resultados de regresión experimentan cambios importantes siempre que se omite una variable explicativa. Por desgracia, ninguna de tales condiciones es necesaria o suficiente para la colinealidad; y además, ninguna proporciona sugerencias útiles, como el tipo de información adicional que se requeriría para resolver el problema de la estimación que afrontan."†

*d*) ". . . toda regresión de serie de tiempo con más de cuatro variables independientes resulta ser una basura."‡

## Ejercicios empíricos

10.26. Klein y Goldberger intentaron ajustar el siguiente modelo de regresión a la economía de Estados Unidos:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

donde $Y$ = consumo, $X_2$ = ingreso salarial, $X_3$ = ingreso no salarial, no procedente del campo, y $X_4$ = ingreso procedente del campo. Pero, como se espera que $X_2$, $X_3$ y $X_4$ sean muy colineales, obtuvieron las siguientes estimaciones de $\beta_3$ y $\beta_4$ del análisis de corte transversal:

* Samprit Chatterjee, Ali S. Hadi y Betram Price, *Regression Analysis by Example*, 3a. ed., John Wiley & Sons, Nueva York, 2000, p. 226.

** Russel Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, p. 186.

† Peter Kennedy, *A Guide to Econometrics*, 4a. ed., MIT Press, Cambridge, Massachusetts, 1998, p. 187.

‡ Esta cita se atribuye al fallecido econometrista Zvi Griliches y se tomó de Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison Wesley, Reading, Massachusetts, 1991, p. 224.

**TABLA 10.12**

Fuente: L. R. Klein y A. S. Goldberger, *An Economic Model of the United States, 1929-1952*, North Holland Publishing Company, Ámsterdam, 1964, p. 131.

| Año | $Y$ | $X_2$ | $X_3$ | $X_4$ | Año | $Y$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|---|---|---|---|---|
| 1936 | 62.8 | 43.41 | 17.10 | 3.96 | 1946 | 95.7 | 76.73 | 28.26 | 9.76 |
| 1937 | 65.0 | 46.44 | 18.65 | 5.48 | 1947 | 98.3 | 75.91 | 27.91 | 9.31 |
| 1938 | 63.9 | 44.35 | 17.09 | 4.37 | 1948 | 100.3 | 77.62 | 32.30 | 9.85 |
| 1939 | 67.5 | 47.82 | 19.28 | 4.51 | 1949 | 103.2 | 78.01 | 31.39 | 7.21 |
| 1940 | 71.3 | 51.02 | 23.24 | 4.88 | 1950 | 108.9 | 83.57 | 35.61 | 7.39 |
| 1941 | 76.6 | 58.71 | 28.11 | 6.37 | 1951 | 108.5 | 90.59 | 37.58 | 7.98 |
| 1945* | 86.3 | 87.69 | 30.29 | 8.96 | 1952 | 111.4 | 95.47 | 35.17 | 7.42 |

* Falta la información correspondiente a los años de guerra 1942-1944. Los datos de los demás años están en miles de millones de dólares de 1939.

$\beta_3 = 0.75\beta_2$ y $\beta_4 = 0.625\beta_2$. Con estas estimaciones reformularon su función de consumo de la siguiente manera:

$$Y_i = \beta_1 + \beta_2(X_{2i} + 0.75X_{3i} + 0.625X_{4i}) + u_i = \beta_1 + \beta_2 Z_i + u_i$$

donde $Z_i = X_{2i} + 0.75X_{3i} + 0.625X_{4i}$.

*a*) Ajuste el modelo modificado a los datos de la tabla 10.12 y obtenga estimaciones de $\beta_1$ a $\beta_4$.

*b*) ¿Como interpretaría la variable Z?

10.27. La tabla 10.13 proporciona cifras sobre importaciones, PIB e índice de precios al consumidor (IPC) de Estados Unidos de 1975 a 2005. Se le pide considerar el siguiente modelo:

$$\ln \text{Importaciones}_t = \beta_1 + \beta_2 \ln \text{PIB}_t + \beta_3 \ln \text{IPC}_t + u_t$$

*a*) Estime los parámetros de este modelo con la información de la tabla.

*b*) ¿Sospecha multicolinealidad en los datos?

*c*) Efectúe las siguientes regresiones:

1) $\ln \text{Importaciones}_t = A_1 + A_2 \ln \text{PIB}_t$

2) $\ln \text{Importaciones}_t = B_1 + B_2 \ln \text{IPC}_t$

3) $\ln \text{PIB}_t = C_1 + C_2 \ln \text{IPC}_t$

Con base en estas regresiones, ¿qué puede decir sobre la naturaleza de la multicolinealidad en los datos?

**TABLA 10.13**
**Importaciones de Estados Unidos, PIB e IPC, 1975-2005 (para todos los consumidores urbanos; 1982-1984 = 100, excepto como se indica)**

Fuente: Department of Labor, Bureau of Labor Statistics.

| Año | IPC | PIB | Importaciones | Año | IPC | PIB | Importaciones |
|---|---|---|---|---|---|---|---|
| 1975 | 53.8 | 1 638.3 | 98 185 | 1991 | 136.2 | 5 995.9 | 491 020 |
| 1976 | 56.9 | 1 825.3 | 124 228 | 1992 | 140.3 | 6 337.7 | 536 528 |
| 1977 | 60.6 | 2 030.9 | 151 907 | 1993 | 144.5 | 6 657.4 | 589 394 |
| 1978 | 65.2 | 2 294.7 | 176 002 | 1994 | 148.2 | 7 072.2 | 668 690 |
| 1979 | 72.6 | 2 563.3 | 212 007 | 1995 | 152.4 | 7 397.7 | 749 374 |
| 1980 | 82.4 | 2 789.5 | 249 750 | 1996 | 156.9 | 7 816.9 | 803 113 |
| 1981 | 90.9 | 3 128.4 | 265 067 | 1997 | 160.5 | 8 304.3 | 876 470 |
| 1982 | 96.5 | 3 225.0 | 247 642 | 1998 | 163.0 | 8 747.0 | 917 103 |
| 1983 | 99.6 | 3 536.7 | 268 901 | 1999 | 166.6 | 9 268.4 | 1 029 980 |
| 1984 | 103.9 | 3 933.2 | 332 418 | 2000 | 172.2 | 9 817.0 | 1 224 408 |
| 1985 | 107.6 | 4 220.3 | 338 088 | 2001 | 177.1 | 10 128.0 | 1 145 900 |
| 1986 | 109.6 | 4 462.8 | 368 425 | 2002 | 179.9 | 10 469.6 | 1 164 720 |
| 1987 | 113.6 | 4 739.5 | 409 765 | 2003 | 184.0 | 10 960.8 | 1 260 717 |
| 1988 | 118.3 | 5 103.8 | 447 189 | 2004 | 188.9 | 11 712.5 | 1 472 926 |
| 1989 | 124.0 | 5 484.4 | 477 665 | 2005 | 195.3 | 12 455.8 | 1 677 371 |
| 1990 | 130.7 | 5 803.1 | 498 438 | | | | |

*d*) Suponga que existe multicolinealidad en los datos, pero que $\hat{\beta}_2$ y $\hat{\beta}_3$ son significativos individualmente en el nivel de 5%, y que la prueba global $F$ es también significativa. En este caso, ¿debe preocupar el problema de colinealidad?

10.28. Consulte el ejercicio 7.19 sobre la función de demanda de pollo en Estados Unidos.

*a*) Con el modelo log-lineal o doble log, estime las diversas regresiones auxiliares. ¿Cuántas hay?

*b*) A partir de estas regresiones auxiliares, ¿cómo decide cuáles regresoras son muy colineales? ¿Qué prueba utiliza? Muestre sus cálculos en detalle.

*c*) Si existe colinealidad significativa en los datos, ¿cuál(es) variable(s) eliminaría(n) para reducir la gravedad del problema de colinealidad? Si lo hace, ¿qué problemas econométricos enfrenta?

*d*) ¿Tiene alguna sugerencia diferente a la de eliminar variables para atenuar el problema de colinealidad? Explique.

10.29. La tabla 10.14 proporciona información sobre los automóviles de pasajeros nuevos vendidos en Estados Unidos como función de diversas variables.

*a*) Desarrolle un modelo lineal o log-lineal apropiado para estimar una función de demanda de automóviles en Estados Unidos.

*b*) Si decide incluir todas las regresoras dadas en la tabla como variables explicativas, ¿espera encontrar el problema de multicolinealidad? ¿Por qué?

*c*) Si espera lo anterior, ¿cómo resolvería el problema? Plantee los supuestos claramente y muestre todos los cálculos de manera explícita.

10.30. Para evaluar la factibilidad de un salario anual garantizado (impuesto sobre la renta negativo), Rand Corporation valoró en un estudio la respuesta de la oferta de trabajo (horas

**TABLA 10.14**
**Datos de automóviles de pasajeros**

Fuente: *Business Statistics, 1986*. Suplemento de *Current Survey of Business*, Department of Commerce, Estados Unidos.

| Año | $Y$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| 1971 | 10 227 | 112.0 | 121.3 | 776.8 | 4.89 | 79 367 |
| 1972 | 10 872 | 111.0 | 125.3 | 839.6 | 4.55 | 82 153 |
| 1973 | 11 350 | 111.1 | 133.1 | 949.8 | 7.38 | 85 064 |
| 1974 | 8 775 | 117.5 | 147.7 | 1 038.4 | 8.61 | 86 794 |
| 1975 | 8 539 | 127.6 | 161.2 | 1 142.8 | 6.16 | 85 846 |
| 1976 | 9 994 | 135.7 | 170.5 | 1 252.6 | 5.22 | 88 752 |
| 1977 | 11 046 | 142.9 | 181.5 | 1 379.3 | 5.50 | 92 017 |
| 1978 | 11 164 | 153.8 | 195.3 | 1 551.2 | 7.78 | 96 048 |
| 1979 | 10 559 | 166.0 | 217.7 | 1 729.3 | 10.25 | 98 824 |
| 1980 | 8 979 | 179.3 | 247.0 | 1 918.0 | 11.28 | 99 303 |
| 1981 | 8 535 | 190.2 | 272.3 | 2 127.6 | 13.73 | 100 397 |
| 1982 | 7 980 | 197.6 | 286.6 | 2 261.4 | 11.20 | 99 526 |
| 1983 | 9 179 | 202.6 | 297.4 | 2 428.1 | 8.69 | 100 834 |
| 1984 | 10 394 | 208.5 | 307.6 | 2 670.6 | 9.65 | 105 005 |
| 1985 | 11 039 | 215.2 | 318.5 | 2 841.1 | 7.75 | 107 150 |
| 1986 | 11 450 | 224.4 | 323.4 | 3 022.1 | 6.31 | 109 597 |

$Y$ = automóviles de pasajeros nuevos vendidos (miles), sin ajuste estacional.
$X_2$ = automóviles nuevos, IPC, 1967 = 100, sin ajuste estacional.
$X_3$ = IPC, todos los renglones, todos los consumidores urbanos, 1967 = 100, sin ajuste estacional.
$X_4$ = ingreso personal disponible (IPD), miles de millones de dólares, sin ajustar por variación estacional.
$X_5$ = tasa de interés, porcentaje, colocación directa de valores de la compañía financiera.
$X_6$ = fuerza laboral civil empleada (miles), sin ajustar por variación estacional.

promedio de trabajo) ante un incremento de los salarios por hora.* Los datos de tal estudio se obtuvieron de una muestra nacional de 6 000 familias cuyo jefe (varón) ganaba menos de 15 000 dólares al año. Los datos se dividieron en 39 grupos demográficos para su análisis. Estos datos se proporcionan en la tabla 10.15. En vista de que para cuatro grupos demográficos había datos faltantes respecto de algunas variables, los datos de la tabla se refieren sólo a 35 de esos grupos. Las definiciones de las diversas variables del análisis se dan al final de la tabla.

**TABLA 10.15**
**Horas de trabajo y otros datos para 35 grupos**

Fuente: D. H. Greenberg y M. Kosters, *Income Guarantees and the Working Poor*, The Rand Corporation, R-579-OEO, diciembre de 1970.

| Observación | Horas | Tasa | IAPE | IAPO | IPAN | Valores | Edad | DEP | Escolaridad |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 157 | 2.905 | 1 121 | 291 | 380 | 7 250 | 38.5 | 2.340 | 10.5 |
| 2 | 2 174 | 2.970 | 1 128 | 301 | 398 | 7 744 | 39.3 | 2.335 | 10.5 |
| 3 | 2 062 | 2.350 | 1 214 | 326 | 185 | 3 068 | 40.1 | 2.851 | 8.9 |
| 4 | 2 111 | 2.511 | 1 203 | 49 | 117 | 1 632 | 22.4 | 1.159 | 11.5 |
| 5 | 2 134 | 2.791 | 1 013 | 594 | 730 | 12 710 | 57.7 | 1.229 | 8.8 |
| 6 | 2 185 | 3.040 | 1 135 | 287 | 382 | 7 706 | 38.6 | 2.602 | 10.7 |
| 7 | 2 210 | 3.222 | 1 100 | 295 | 474 | 9 338 | 39.0 | 2.187 | 11.2 |
| 8 | 2 105 | 2.493 | 1 180 | 310 | 255 | 4 730 | 39.9 | 2.616 | 9.3 |
| 9 | 2 267 | 2.838 | 1 298 | 252 | 431 | 8 317 | 38.9 | 2.024 | 11.1 |
| 10 | 2 205 | 2.356 | 885 | 264 | 373 | 6 789 | 38.8 | 2.662 | 9.5 |
| 11 | 2 121 | 2.922 | 1 251 | 328 | 312 | 5 907 | 39.8 | 2.287 | 10.3 |
| 12 | 2 109 | 2.499 | 1 207 | 347 | 271 | 5 069 | 39.7 | 3.193 | 8.9 |
| 13 | 2 108 | 2.796 | 1 036 | 300 | 259 | 4 614 | 38.2 | 2.040 | 9.2 |
| 14 | 2 047 | 2.453 | 1 213 | 297 | 139 | 1 987 | 40.3 | 2.545 | 9.1 |
| 15 | 2 174 | 3.582 | 1 141 | 414 | 498 | 10 239 | 40.0 | 2.064 | 11.7 |
| 16 | 2 067 | 2.909 | 1 805 | 290 | 239 | 4 439 | 39.1 | 2.301 | 10.5 |
| 17 | 2 159 | 2.511 | 1 075 | 289 | 308 | 5 621 | 39.3 | 2.486 | 9.5 |
| 18 | 2 257 | 2.516 | 1 093 | 176 | 392 | 7 293 | 37.9 | 2.042 | 10.1 |
| 19 | 1 985 | 1.423 | 553 | 381 | 146 | 1 866 | 40.6 | 3.833 | 6.6 |
| 20 | 2 184 | 3.636 | 1 091 | 291 | 560 | 11 240 | 39.1 | 2.328 | 11.6 |
| 21 | 2 084 | 2.983 | 1 327 | 331 | 296 | 5 653 | 39.8 | 2.208 | 10.2 |
| 22 | 2 051 | 2.573 | 1 194 | 279 | 172 | 2 806 | 40.0 | 2.362 | 9.1 |
| 23 | 2 127 | 3.262 | 1 226 | 314 | 408 | 8 042 | 39.5 | 2.259 | 10.8 |
| 24 | 2 102 | 3.234 | 1 188 | 414 | 352 | 7 557 | 39.8 | 2.019 | 10.7 |
| 25 | 2 098 | 2.280 | 973 | 364 | 272 | 4 400 | 40.6 | 2.661 | 8.4 |
| 26 | 2 042 | 2.304 | 1 085 | 328 | 140 | 1 739 | 41.8 | 2.444 | 8.2 |
| 27 | 2 181 | 2.912 | 1 072 | 304 | 383 | 7 340 | 39.0 | 2.337 | 10.2 |
| 28 | 2 186 | 3.015 | 1 122 | 30 | 352 | 7 292 | 37.2 | 2.046 | 10.9 |
| 29 | 2 188 | 3.010 | 990 | 366 | 374 | 7 325 | 38.4 | 2.847 | 10.6 |
| 30 | 2 077 | 1.901 | 350 | 209 | 95 | 1 370 | 37.4 | 4.158 | 8.2 |
| 31 | 2 196 | 3.009 | 947 | 294 | 342 | 6 888 | 37.5 | 3.047 | 10.6 |
| 32 | 2 093 | 1.899 | 342 | 311 | 120 | 1 425 | 37.5 | 4.512 | 8.1 |
| 33 | 2 173 | 2.959 | 1 116 | 296 | 387 | 7 625 | 39.2 | 2.342 | 10.5 |
| 34 | 2 179 | 2.971 | 1 128 | 312 | 397 | 7 779 | 39.4 | 2.341 | 10.5 |
| 35 | 2 200 | 2.980 | 1 126 | 204 | 393 | 7 885 | 39.2 | 2.341 | 10.6 |

*Notas:* Horas = horas promedio trabajadas durante el año.
Tasa = salario promedio por hora (dólares).
IAPE = ingresos anuales promedio de la esposa (dólares).
IAPO = ingresos anuales promedio de otros miembros de la familia (dólares).
IPAN = ingreso promedio anual no devengado.
Valores = bienes familiares promedio (cuentas bancarias, etc.) (dólares).
Edad = edad promedio del entrevistado.
DEP = número promedio de dependientes.
Escolaridad = nivel máximo de escolaridad promedio completado.

* D.H. Greenberg y M. Kosters, *Income Guarantees and the Working Poor*, Rand Corporation, R-579-OEO, diciembre de 1970.

*a*) Realice la regresión de las horas promedio trabajadas durante un año sobre las variables suministradas en la tabla e interprete su regresión.

*b*) ¿Existe evidencia de multicolinealidad en los datos? ¿Cómo sabe?

*c*) Calcule las medidas del factor inflacionario de la varianza (VFI) y de la TOL para las diversas regresoras.

*d*) Si existe un problema de multicolinealidad, ¿qué acciones correctivas, si acaso hay alguna, tomaría?

*e*) ¿Qué sugiere este estudio sobre la viabilidad de un impuesto sobre la renta negativo?

10.31. La tabla 10.16 muestra datos sobre la tasa de criminalidad en 47 estados de EU durante 1960. Intente elaborar un modelo apropiado para explicar la tasa de criminalidad respecto de las 14 variables socioeconómicas de la tabla. Ponga especial atención al problema de la colinealidad cuando desarrolle su modelo.

10.32. Consulte los datos Longley suministrados en la sección 10.10. Repita la regresión dada en esa tabla pero sin los datos de 1962; es decir, haga la regresión del periodo 1947-1961. Compare ambas regresiones. ¿Qué conclusión general deduce de este ejercicio?

10.33. *Datos Longley actualizados*. Ampliamos los datos de la sección 10.10 para incluir observaciones de 1959-2005. Los nuevos datos aparecen en la tabla 10.17. Los datos se relacionan con $Y$ = número de personas empleadas, en millares; $X_1$ = deflactor de precios implícito del PNB; $X_2$ = PNB, millones de dólares; $X_3$ = número de personas desempleadas, en millares; $X_4$ = número de personas en las fuerzas armadas, en millares; $X_5$ = población no institucionalizada mayor de 16 años, y $X_6$ = año, igual a 1 en 1959, 2 en 1960 y 47 en 2005.

*a*) Trace diagramas de dispersión, como se indica en el capítulo, para evaluar las relaciones entre las variables independientes. ¿Hay relaciones fuertes? ¿Parecen lineales?

*b*) Elabore una matriz de correlación. ¿Qué variables parecen relacionarse más ente sí, sin incluir la dependiente?

*c*) Ejecute una regresión estándar de MCO para pronosticar el número de personas empleadas en millares. ¿Los coeficientes de las variables independientes se comportan como esperaría?

*d*) Con base en los resultados anteriores, ¿cree que estos datos sufren de multicolinealidad?

*10.34. Conforme el queso se añeja, varios procesos químicos tienen lugar y determinan el sabor del producto final. Los datos presentados de la tabla 10.18 corresponden a concentraciones de varias sustancias químicas en una muestra de 30 quesos cheddar maduros, y medidas subjetivas respecto del sabor de cada muestra. Las variables acético y $H_2S$ corresponden al logaritmo natural de la concentración de ácido acético y sulfuro de hidrógeno, respectivamente. La variable láctico no se transformó logarítmicamente.

*a*) Trace un diagrama de dispersión de las cuatro variables.

*b*) Ejecute una regresión bivariada del sabor sobre acético y $H_2S$, e interprete los resultados.

*c*) Ejecute una regresión bivariada de sabor sobre láctico y $H_2S$, e interprete los resultados.

*d*) Ejecute una regresión múltiple de sabor sobre acético, $H_2S$ y láctico. Interprete los resultados.

*e*) Con lo que ahora sabe sobre multicolinealidad, ¿cómo elegiría entre estas regresiones?

*f*) ¿Qué conclusiones generales deduce de su análisis?

* Opcional.

**TABLA 10.16 Datos sobre criminalidad en 47 estados de Estados Unidos en 1960**

| Observación | $R$ | Edad | $S$ | EP | $GX_0$ | $GX_1$ | FT | $H$ | $N$ | NB | $U_1$ | $U_2$ | $W$ | $X$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 79.1 | 151 | 1 | 91 | 58 | 56 | 510 | 950 | 33 | 301 | 108 | 41 | 394 | 261 |
| 2 | 163.5 | 143 | 0 | 113 | 103 | 95 | 583 | 1 012 | 13 | 102 | 96 | 36 | 557 | 194 |
| 3 | 57.8 | 142 | 1 | 89 | 45 | 44 | 533 | 969 | 18 | 219 | 94 | 33 | 318 | 250 |
| 4 | 196.9 | 136 | 0 | 121 | 149 | 141 | 577 | 994 | 157 | 80 | 102 | 39 | 673 | 167 |
| 5 | 123.4 | 141 | 0 | 121 | 109 | 101 | 591 | 985 | 18 | 30 | 91 | 20 | 578 | 174 |
| 6 | 68.2 | 121 | 0 | 110 | 118 | 115 | 547 | 964 | 25 | 44 | 84 | 29 | 689 | 126 |
| 7 | 96.3 | 127 | 1 | 111 | 82 | 79 | 519 | 982 | 4 | 139 | 97 | 38 | 620 | 168 |
| 8 | 155.5 | 131 | 1 | 109 | 115 | 109 | 542 | 969 | 50 | 179 | 79 | 35 | 472 | 206 |
| 9 | 85.6 | 157 | 1 | 90 | 65 | 62 | 553 | 955 | 39 | 286 | 81 | 28 | 421 | 239 |
| 10 | 70.5 | 140 | 0 | 118 | 71 | 68 | 632 | 1 029 | 7 | 15 | 100 | 24 | 526 | 174 |
| 11 | 167.4 | 124 | 0 | 105 | 121 | 116 | 580 | 966 | 101 | 106 | 77 | 35 | 657 | 170 |
| 12 | 84.9 | 134 | 0 | 108 | 75 | 71 | 595 | 972 | 47 | 59 | 83 | 31 | 580 | 172 |
| 13 | 51.1 | 128 | 0 | 113 | 67 | 60 | 624 | 972 | 28 | 10 | 77 | 25 | 507 | 206 |
| 14 | 66.4 | 135 | 0 | 117 | 62 | 61 | 595 | 986 | 22 | 46 | 77 | 27 | 529 | 190 |
| 15 | 79.8 | 152 | 1 | 87 | 57 | 53 | 530 | 986 | 30 | 72 | 92 | 43 | 405 | 264 |
| 16 | 94.6 | 142 | 1 | 88 | 81 | 77 | 497 | 956 | 33 | 321 | 116 | 47 | 427 | 247 |
| 17 | 53.9 | 143 | 0 | 110 | 66 | 63 | 537 | 977 | 10 | 6 | 114 | 35 | 487 | 166 |
| 18 | 92.9 | 135 | 1 | 104 | 123 | 115 | 537 | 978 | 31 | 170 | 89 | 34 | 631 | 165 |
| 19 | 75.0 | 130 | 0 | 116 | 128 | 128 | 536 | 934 | 51 | 24 | 78 | 34 | 627 | 135 |
| 20 | 122.5 | 125 | 0 | 108 | 113 | 105 | 567 | 985 | 78 | 94 | 130 | 58 | 626 | 166 |
| 21 | 74.2 | 126 | 0 | 108 | 74 | 67 | 602 | 984 | 34 | 12 | 102 | 33 | 557 | 195 |
| 22 | 43.9 | 157 | 1 | 89 | 47 | 44 | 512 | 962 | 22 | 423 | 97 | 34 | 288 | 276 |
| 23 | 121.6 | 132 | 0 | 96 | 87 | 83 | 564 | 953 | 43 | 92 | 83 | 32 | 513 | 227 |
| 24 | 96.8 | 131 | 0 | 116 | 78 | 73 | 574 | 1 038 | 7 | 36 | 142 | 42 | 540 | 176 |
| 25 | 52.3 | 130 | 0 | 116 | 63 | 57 | 641 | 984 | 14 | 26 | 70 | 21 | 486 | 196 |
| 26 | 199.3 | 131 | 0 | 121 | 160 | 143 | 631 | 1 071 | 3 | 77 | 102 | 41 | 674 | 152 |
| 27 | 34.2 | 135 | 0 | 109 | 69 | 71 | 540 | 965 | 6 | 4 | 80 | 22 | 564 | 139 |
| 28 | 121.6 | 152 | 0 | 112 | 82 | 76 | 571 | 1 018 | 10 | 79 | 103 | 28 | 537 | 215 |
| 29 | 104.3 | 119 | 0 | 107 | 166 | 157 | 521 | 938 | 168 | 89 | 92 | 36 | 637 | 154 |
| 30 | 69.6 | 166 | 1 | 89 | 58 | 54 | 521 | 973 | 46 | 254 | 72 | 26 | 396 | 237 |
| 31 | 37.3 | 140 | 0 | 93 | 55 | 54 | 535 | 1 045 | 6 | 20 | 135 | 40 | 453 | 200 |
| 32 | 75.4 | 125 | 0 | 109 | 90 | 81 | 586 | 964 | 97 | 82 | 105 | 43 | 617 | 163 |
| 33 | 107.2 | 147 | 1 | 104 | 63 | 64 | 560 | 972 | 23 | 95 | 76 | 24 | 462 | 233 |
| 34 | 92.3 | 126 | 0 | 118 | 97 | 97 | 542 | 990 | 18 | 21 | 102 | 35 | 589 | 166 |
| 35 | 65.3 | 123 | 0 | 102 | 97 | 87 | 526 | 948 | 113 | 76 | 124 | 50 | 572 | 158 |
| 36 | 127.2 | 150 | 0 | 100 | 109 | 98 | 531 | 964 | 9 | 24 | 87 | 38 | 559 | 153 |
| 37 | 83.1 | 177 | 1 | 87 | 58 | 56 | 638 | 974 | 24 | 349 | 76 | 28 | 382 | 254 |
| 38 | 56.6 | 133 | 0 | 104 | 51 | 47 | 599 | 1 024 | 7 | 40 | 99 | 27 | 425 | 225 |
| 39 | 82.6 | 149 | 1 | 88 | 61 | 54 | 515 | 953 | 36 | 165 | 86 | 35 | 395 | 251 |
| 40 | 115.1 | 145 | 1 | 104 | 82 | 74 | 560 | 981 | 96 | 126 | 88 | 31 | 488 | 228 |
| 41 | 88.0 | 148 | 0 | 122 | 72 | 66 | 601 | 998 | 9 | 19 | 84 | 20 | 590 | 144 |
| 42 | 54.2 | 141 | 0 | 109 | 56 | 54 | 523 | 968 | 4 | 2 | 107 | 37 | 489 | 170 |
| 43 | 82.3 | 162 | 1 | 99 | 75 | 70 | 522 | 996 | 40 | 208 | 73 | 27 | 496 | 224 |
| 44 | 103.0 | 136 | 0 | 121 | 95 | 96 | 574 | 1 012 | 29 | 36 | 111 | 37 | 622 | 162 |
| 45 | 45.5 | 139 | 1 | 88 | 46 | 41 | 480 | 968 | 19 | 49 | 135 | 53 | 457 | 249 |
| 46 | 50.8 | 126 | 0 | 104 | 106 | 97 | 599 | 989 | 40 | 24 | 78 | 25 | 593 | 171 |
| 47 | 84.9 | 130 | 0 | 121 | 90 | 91 | 623 | 1 049 | 3 | 22 | 113 | 40 | 588 | 160 |

Fuente: W. Vandaele, "Participation in Illegitimate Activities: Erlich Revisted", en A. Blumstein, J. Cohen y Nagin D. (eds.), *Deterrence and Incapacitation*, National Academy of Sciences, 1978, pp. 270-335.

Definición de las variables:

$R$ = tasa de criminalidad, número de delitos reportados a la policía por millón de personas.
Edad = número de hombres entre 14 y 24 años de edad por cada 1 000.
$S$ = variable indicadora de los estados sureños (1 = sí, 0 = no).
EP = número promedio de años de escolaridad multiplicado por 10 para personas de 25 años o mayores.
$GX_0$ = gastos de la policía *per cápita* a cargo del estado y el gobierno local en 1960.
$GX_1$ = gastos de la policía *per cápita* a cargo del estado y el gobierno local en 1959.
FT = tasa de participación de la fuerza de trabajo por cada 1 000 hombres civiles urbanos entre 14 y 24 años de edad.
$H$ = número de hombres por cada 1 000 mujeres.
$N$ = tamaño de la población estatal en cientos de miles.
NB = número de no blancos por cada 1 000 personas.
$D_1$ = tasa de desempleo de hombres urbanos por cada 100 habitantes de entre 14 y 24 años de edad.
$D_2$ = tasa de desempleo de hombres urbanos por cada 100 habitantes de entre 35 y 39 años de edad.
$W$ = valor medio de bienes y valores transferibles, o ingreso familiar en decenas de dólares.
$X$ = número de familias por 1 000 salarios iguales a la ½ de la mediana del ingreso.
Observación = estado (47 estados para 1960).

**TABLA 10.17**
**Datos Longley actualizados, 1959-2005**

Fuente: Department of Labor, Bureau of Labor Statistics, Estados Unidos, y http://diadapp.dmdc.osd.mil/personnel/MILITARY/Miltop.htm.

| Observación | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|---|
| 1959 | 64 630 | 82.908 | 509 300 | 3 740 | 2 552 | 120 287 | 1 |
| 1960 | 65 778 | 84.074 | 529 500 | 3 852 | 2 514 | 121 836 | 2 |
| 1961 | 65 746 | 85.015 | 548 200 | 4 714 | 2 573 | 123 404 | 3 |
| 1962 | 66 702 | 86.186 | 589 700 | 3 911 | 2 827 | 124 864 | 4 |
| 1963 | 67 762 | 87.103 | 622 200 | 4 070 | 2 737 | 127 274 | 5 |
| 1964 | 69 305 | 88.438 | 668 500 | 3 786 | 2 738 | 129 427 | 6 |
| 1965 | 71 088 | 90.055 | 724 400 | 3 366 | 2 722 | 131 541 | 7 |
| 1966 | 72 895 | 92.624 | 792 900 | 2 875 | 3 123 | 133 650 | 8 |
| 1967 | 74 372 | 95.491 | 838 000 | 2 975 | 3 446 | 135 905 | 9 |
| 1968 | 75 920 | 99.56 | 916 100 | 2 817 | 3 535 | 138 171 | 10 |
| 1969 | 77 902 | 104.504 | 990 700 | 2 832 | 3 506 | 140 461 | 11 |
| 1970 | 78 678 | 110.046 | 1 044 900 | 4 093 | 3 188 | 143 070 | 12 |
| 1971 | 79 367 | 115.549 | 1 134 700 | 5 016 | 2 816 | 145 826 | 13 |
| 1972 | 82 153 | 120.556 | 1 246 800 | 4 882 | 2 449 | 148 592 | 14 |
| 1973 | 85 064 | 127.307 | 1 395 300 | 4 365 | 2 327 | 151 476 | 15 |
| 1974 | 86 794 | 138.82 | 1 515 500 | 5 156 | 2 229 | 154 378 | 16 |
| 1975 | 85 846 | 151.857 | 1 651 300 | 7 929 | 2 180 | 157 344 | 17 |
| 1976 | 88 752 | 160.68 | 1 842 100 | 7 406 | 2 144 | 160 319 | 18 |
| 1977 | 92 017 | 170.884 | 2 051 200 | 6 991 | 2 133 | 163 377 | 19 |
| 1978 | 96 048 | 182.863 | 2 316 300 | 6 202 | 2 117 | 166 422 | 20 |
| 1979 | 98 824 | 198.077 | 2 595 300 | 6 137 | 2 088 | 169 440 | 21 |
| 1980 | 99 303 | 216.073 | 2 823 700 | 7 637 | 2 102 | 172 437 | 22 |
| 1981 | 100 397 | 236.385 | 3 161 400 | 8 273 | 2 142 | 174 929 | 23 |
| 1982 | 99 526 | 250.798 | 3 291 500 | 10 678 | 2 179 | 177 176 | 24 |
| 1983 | 100 834 | 260.68 | 3 573 800 | 10 717 | 2 199 | 179 234 | 25 |
| 1984 | 105 005 | 270.496 | 3 969 500 | 8 539 | 2 219 | 181 192 | 26 |
| 1985 | 107 150 | 278.759 | 4 246 800 | 8 312 | 2 234 | 183 174 | 27 |
| 1986 | 109 597 | 284.895 | 4 480 600 | 8 237 | 2 244 | 185 284 | 28 |
| 1987 | 112 440 | 292.691 | 4 757 400 | 7 425 | 2 257 | 187 419 | 29 |
| 1988 | 114 968 | 302.68 | 5 127 400 | 6 701 | 2 224 | 189 233 | 30 |
| 1989 | 117 342 | 314.179 | 5 510 600 | 6 528 | 2 208 | 190 862 | 31 |
| 1990 | 118 793 | 326.357 | 5 837 900 | 7 047 | 2 167 | 192 644 | 32 |
| 1991 | 117 718 | 337.747 | 6 026 300 | 8 628 | 2 118 | 194 936 | 33 |
| 1992 | 118 492 | 345.477 | 6 367 400 | 9 613 | 1 966 | 197 205 | 34 |
| 1993 | 120 259 | 353.516 | 6 689 300 | 8 940 | 1 760 | 199 622 | 35 |
| 1994 | 123 060 | 361.026 | 7 098 400 | 7 996 | 1 673 | 201 970 | 36 |
| 1995 | 124 900 | 368.444 | 7 433 400 | 7 404 | 1 579 | 204 420 | 37 |
| 1996 | 126 708 | 375.429 | 7 851 900 | 7 236 | 1 502 | 207 087 | 38 |
| 1997 | 129 558 | 381.663 | 8 337 300 | 6 739 | 1 457 | 209 846 | 39 |
| 1998 | 131 463 | 385.881 | 8 768 300 | 6 210 | 1 423 | 212 638 | 40 |
| 1999 | 133 488 | 391.452 | 9 302 200 | 5 880 | 1 380 | 215 404 | 41 |
| 2000 | 136 891 | 399.986 | 9 855 900 | 5 692 | 1 405 | 218 061 | 42 |
| 2001 | 136 933 | 409.582 | 10 171 600 | 6 801 | 1 412 | 220 800 | 43 |
| 2002 | 136 485 | 416.704 | 10 500 200 | 8 378 | 1 425 | 223 532 | 44 |
| 2003 | 137 736 | 425.553 | 11 017 600 | 8 774 | 1 423 | 226 223 | 45 |
| 2004 | 139 252 | 437.795 | 11 762 100 | 8 149 | 1 411 | 228 892 | 46 |
| 2005 | 141 730 | 451.946 | 12 502 400 | 7 591 | 1 378 | 231 552 | 47 |

**TABLA 10.18**
**Sustancias químicas en quesos**

Fuente: http://lib.stat.cmu.edu/DASL/Datafiles/Cheese.html.

| Obs. | Sabor | Acético | $H_2S$ | Láctico |
|---|---|---|---|---|
| 1 | 12.30000 | 4.543000 | 3.135000 | 0.860000 |
| 2 | 20.90000 | 5.159000 | 5.043000 | 1.530000 |
| 3 | 39.00000 | 5.366000 | 5.438000 | 1.570000 |
| 4 | 47.90000 | 5.759000 | 7.496000 | 1.810000 |
| 5 | 5.600000 | 4.663000 | 3.807000 | 0.990000 |
| 6 | 25.90000 | 5.697000 | 7.601000 | 1.090000 |
| 7 | 37.30000 | 5.892000 | 8.726000 | 1.290000 |
| 8 | 21.90000 | 6.078000 | 7.966000 | 1.780000 |
| 9 | 18.10000 | 4.898000 | 3.850000 | 1.290000 |
| 10 | 21.00000 | 5.242000 | 4.174000 | 1.580000 |
| 11 | 34.90000 | 5.740000 | 6.142000 | 1.680000 |
| 12 | 57.20000 | 6.446000 | 7.908000 | 1.900000 |
| 13 | 0.700000 | 4.477000 | 2.996000 | 1.060000 |
| 14 | 25.90000 | 5.236000 | 4.942000 | 1.300000 |
| 15 | 54.90000 | 6.151000 | 6.752000 | 1.520000 |
| 16 | 40.90000 | 3.365000 | 9.588000 | 1.740000 |
| 17 | 15.90000 | 4.787000 | 3.912000 | 1.160000 |
| 18 | 6.400000 | 5.142000 | 4.700000 | 1.490000 |
| 19 | 18.00000 | 5.247000 | 6.174000 | 1.630000 |
| 20 | 38.90000 | 5.438000 | 9.064000 | 1.990000 |
| 21 | 14.00000 | 4.564000 | 4.949000 | 1.150000 |
| 22 | 15.20000 | 5.298000 | 5.220000 | 1.330000 |
| 23 | 32.00000 | 5.455000 | 9.242000 | 1.440000 |
| 24 | 56.70000 | 5.855000 | 10.19900 | 2.010000 |
| 25 | 16.80000 | 5.366000 | 3.664000 | 1.310000 |
| 26 | 11.60000 | 6.043000 | 3.219000 | 1.460000 |
| 27 | 26.50000 | 6.458000 | 6.962000 | 1.720000 |
| 28 | 0.700000 | 5.328000 | 3.912000 | 1.250000 |
| 29 | 13.40000 | 5.802000 | 6.685000 | 1.080000 |
| 30 | 5.500000 | 6.176000 | 4.787000 | 1.250000 |

Capítulo 11

# Heteroscedasticidad: ¿qué pasa si la varianza del error no es constante?

Un supuesto importante del modelo clásico de regresión lineal (supuesto 4) es que las perturbaciones $u_i$ que aparecen en la función de regresión poblacional son homoscedásticas; es decir, que todas tienen la misma varianza. En este capítulo examinaremos la validez de este supuesto y también lo que sucede si no se cumple. Como en el capítulo 10, buscamos respuestas a las siguientes preguntas:

1. ¿Cuál es la naturaleza de la heteroscedasticidad?
2. ¿Cuáles son sus consecuencias?
3. ¿Cómo se detecta?
4. ¿Qué remedios existen?

## 11.1 Naturaleza de la heteroscedasticidad

Como vimos en el capítulo 3, un supuesto importante del modelo clásico de regresión lineal es que la varianza de cada término de perturbación $u_i$, condicional a los valores seleccionados de las variables explicativas, es algún número constante igual a $\sigma^2$. Éste es el supuesto de **homoscedasticidad**, o *igual* (homo) *dispersión* (cedasticidad), es decir, *igual varianza*. Simbólicamente,

$$E\left(u_i^2\right) = \sigma^2 \qquad i = 1, 2, \ldots, n \tag{11.1.1}$$

Gráficamente, la homoscedasticidad en el modelo de regresión con dos variables se observa en la figura 3.4, la cual, por conveniencia, se reproduce como la figura 11.1. Como lo indica esta figura, la varianza condicional de $Y_i$ (la cual es igual a la de $u_i$), condicional a las $X_i$ dadas, permanece igual sin importar los valores que tome la variable $X$.

En contraste, observe la figura 11.2, donde la varianza condicional de $Y_i$ aumenta a medida que lo hace $X$. Aquí, las varianzas de $Y_i$ no son las mismas. Por tanto, hay heteroscedasticidad. Simbólicamente,

$$E\left(u_i^2\right) = \sigma_i^2 \tag{11.1.2}$$

**FIGURA 11.1**
Perturbaciones homoscedásticas.

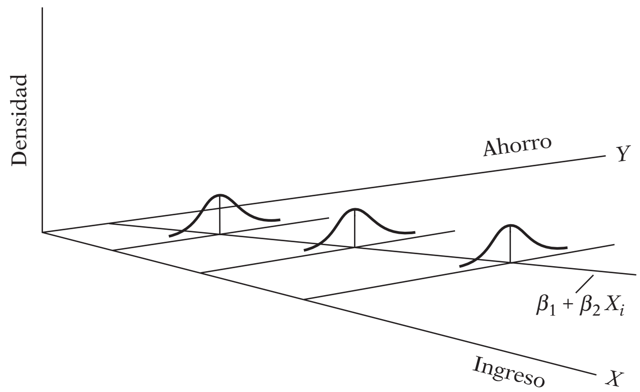
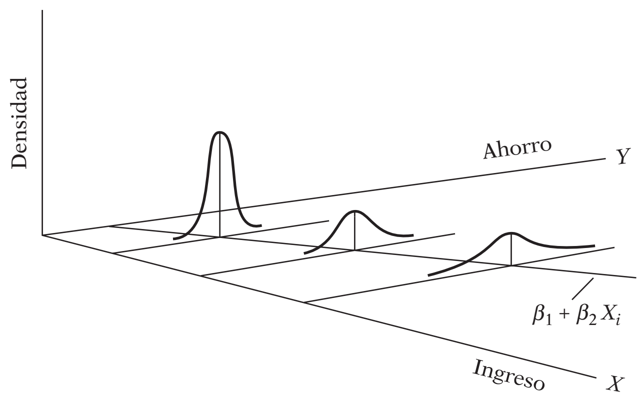



**FIGURA 11.2**
Perturbaciones heteroscedásticas.




Observe el subíndice de $\sigma^2$, que indica que las varianzas condicionales de $u_i$ (= varianzas condicionales de $Y_i$) ya no son constantes.

Para entender la diferencia entre homoscedasticidad y heteroscedasticidad, suponga que en el modelo con dos variables $Y_i = \beta_1 + \beta_2 X_i + u_i$, $Y$ representa el ahorro y $X$ el ingreso. Las figuras 11.1 y 11.2 indican que, a medida que el ingreso aumenta, el ahorro en promedio también aumenta. Pero, en la figura 11.1, la varianza del ahorro permanece igual en todos los niveles de ingreso, mientras que en la figura 11.2 se incrementa con aumentos del ingreso. Parece que en la figura 11.2, en promedio, las familias de ingresos más altos ahorran más que las de ingresos más bajos, pero también hay más variabilidad en su ahorro.

Hay diversas razones por las cuales las varianzas de $u_i$ pueden ser variables, algunas de las cuales son las siguientes.[1]

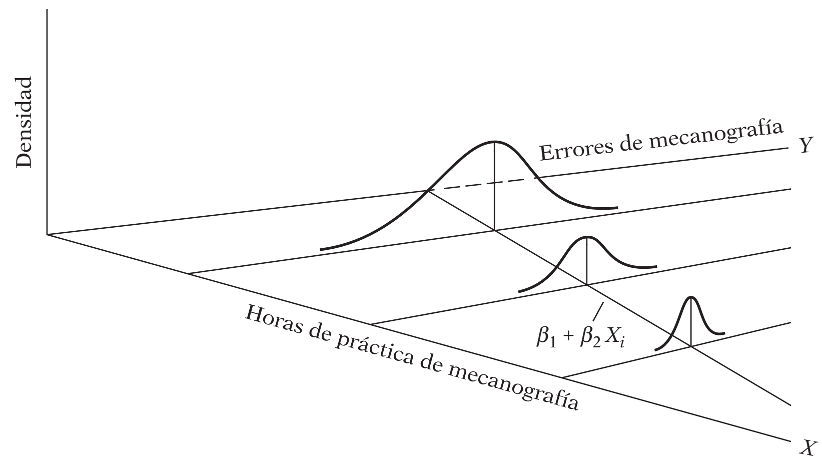
1. Con base en los *modelos de aprendizaje de los errores*, a medida que la gente aprende, disminuyen sus errores de comportamiento con el tiempo. En este caso, esperamos que $\sigma_i^2$ se reduzca. Como ejemplo, considere la figura 11.3, que relaciona el número de errores cometidos en una prueba de mecanografía durante un periodo dado. Como indica la figura 11.3, a medida que aumenta el número de horas de esta práctica, se reduce el número promedio de errores de mecanografía, al igual que sus varianzas.

2. A medida que aumentan los ingresos, la gente posee más *ingreso discrecional*[2] y, por tanto, tiene mayores posibilidades de decidir cómo disponer de su ingreso. En consecuencia,

[1] Véase Stefan Valavanis, *Econometrics,* McGraw-Hill, Nueva York, 1959, p. 48.

[2] Como señala Valavanis, "El ingreso aumenta y ahora la gente escasamente distingue los dólares, mientras que antes distinguía las monedas de diez centavos", *ibid.*, p. 48.

**FIGURA 11.3**
Ilustración de la heteroscedasticidad.




es probable que $\sigma_i^2$ aumente con el ingreso. Así, en la regresión del ahorro sobre el ingreso, es probable encontrar que $\sigma_i^2$ aumenta con el ingreso (como en la figura 11.2), pues las personas tienen mayores posibilidades de determinar su comportamiento respecto del ahorro. En forma similar, se espera que las compañías con mayores ganancias presenten mayor variabilidad en sus políticas de dividendos que las compañías cuyas ganancias son menores. Además, es probable que las empresas *orientadas al crecimiento* presenten una mayor variabilidad en sus tasas de pago de dividendos que las empresas ya establecidas.

3. A medida que mejoran las técnicas de recolección de datos, es probable que $\sigma_i^2$ se reduzca. Así, es probable que los bancos con equipos complejos de procesamiento de información cometan menos errores en los informes mensuales o trimestrales de sus clientes que los bancos que no los posean.

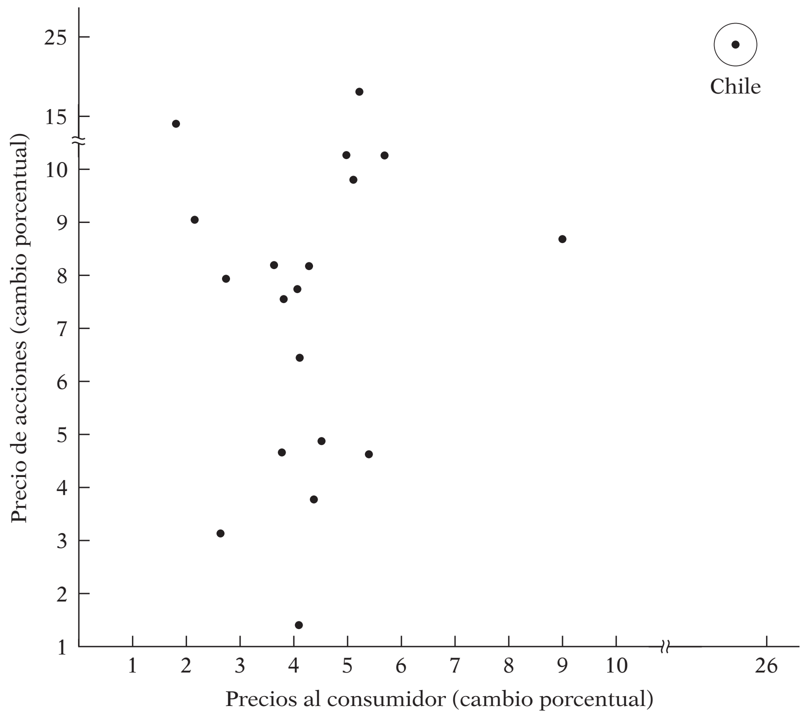
4. La heteroscedasticidad también surge por la presencia de **datos atípicos** o **aberrantes**. Una observación atípica es la que es muy diferente (muy pequeña o muy grande) en relación con las demás observaciones en la muestra. De manera más precisa, un dato atípico es una observación que proviene de una población distinta a la que genera las demás observaciones de la muestra.[3] La inclusión o exclusión de una observación de este tipo, en especial si el tamaño de la muestra es pequeño, puede alterar sustancialmente los resultados del análisis de regresión.

Como ejemplo considere el diagrama de dispersión de la figura 11.4. Con base en los datos de la tabla 11.9 del ejercicio 11.22, en esta figura graficamos la tasa de cambio porcentual de los precios de las acciones ($Y$) y los precios al consumidor ($X$) para el periodo posterior a la Segunda Guerra Mundial hasta 1969 en 20 países. En esta figura, la observación sobre $Y$ y $X$ para Chile se considera atípica, pues los valores $Y$ y $X$ son mucho más grandes que para el resto de los países. En estas situaciones es difícil mantener el supuesto de homoscedasticidad. En el ejercicio 11.22 se le pide ver qué sucede con los resultados de la regresión si se retiran del análisis las observaciones de Chile.

5. Otra fuente de heteroscedasticidad surge de la violación del supuesto 9 del MCRL, que establece que el modelo de regresión está correctamente especificado. Aunque analizaremos más a fondo los errores de especificación en el capítulo 13, con mucha frecuencia lo que parece heteroscedasticidad puede deberse a que se omitan del modelo algunas variables importantes. Así, en la función de demanda de un bien, si no se incluyen los precios de los bienes que le son complementarios o con los que compite (sesgo de variable omitida), los residuos de la regresión pueden dar la clara impresión de que la varianza del error no es constante. Pero si se incluyen en el modelo las variables omitidas, esa impresión puede desaparecer.

[3] El autor agradece a Michael McAleer por este señalamiento.

**FIGURA 11.4**
Relación entre precios de acciones y precios al consumidor.

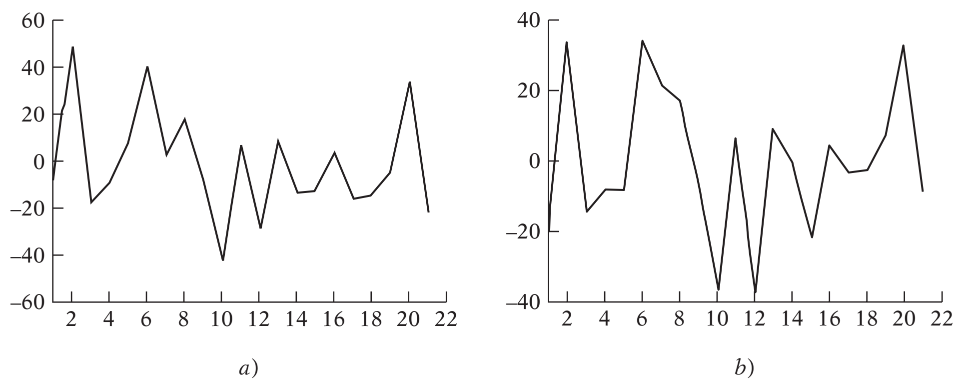



Como un ejemplo concreto, recuerde el estudio de los impactos publicitarios retenidos ($Y$) en relación con el gasto en publicidad ($X$). (Consulte el ejercicio 8.32.) Si sólo hace la regresión de $Y$ sobre $X$ y se observan los residuos de esta regresión, se presenta un patrón, pero si se efectúa la regresión de $Y$ sobre $X$ y sobre $X^2$, se verá otro patrón, lo cual se distingue con claridad en la figura 11.5. Ya comprobamos que $X^2$ pertenece al modelo. (Consulte el ejercicio 8.32.)

6. Otra fuente de la heteroscedasticidad es la **asimetría** en la distribución de una o más regresoras incluidas en el modelo. Los ejemplos los constituyen variables económicas como ingreso, riqueza y escolaridad. Es bien sabido que la distribución del ingreso y la riqueza en la mayoría de las sociedades es desigual, pues la mayor parte del ingreso o la riqueza le corresponde a unos cuantos individuos pertenecientes a los estratos superiores.

7. Otras fuentes de heteroscedasticidad: como señala David Hendry, la heteroscedasticidad también surge debido a 1) la incorrecta transformación de los datos (por ejemplo, las transformaciones de razón o de primeras diferencias) y 2) una forma funcional incorrecta (por ejemplo, modelos lineales frente a modelos log-lineales).[4]

**FIGURA 11.5**
Residuos de las regresiones de *a*) impactos publicitarios en el gasto publicitario (GASP) y *b*) impactos publicitarios en GASP y GASP².




[4] David F. Hendry, *Dynamic Econometrics,* Oxford University Press, 1995, p. 45.

Observe que el problema de heteroscedasticidad es quizá más común en la información de corte transversal que en la de series de tiempo. En la información de corte transversal, por lo general se trata con miembros de una población en un momento dado, como consumidores individuales o sus familias, empresas, industrias o subdivisiones geográficas, como estados, países, ciudades, etc. Además, estos miembros pueden ser de diferentes tamaños, como empresas pequeñas, medianas o grandes, o ingresos bajos, medios o altos. En las series de tiempo, por el contrario, las variables tienden a ser de órdenes de magnitud similares porque suele recopilarse información sobre el mismo fenómeno o hecho a lo largo de un periodo. Son ejemplos el PIB, el gasto de consumo, el ahorro o el empleo en Estados Unidos, por decir, de 1955 a 2005.

A manera de ilustración sobre la heteroscedasticidad que es posible encontrar en un análisis de corte transversal, considere la tabla 11.1. Esta tabla presenta información sobre salarios por empleado en 10 industrias manufactureras de bienes no duraderos, clasificadas según el número de empleados de la empresa o establecimiento en 1958. En la tabla se presentan además cifras de productividad promedio de nueve clases de empleados.

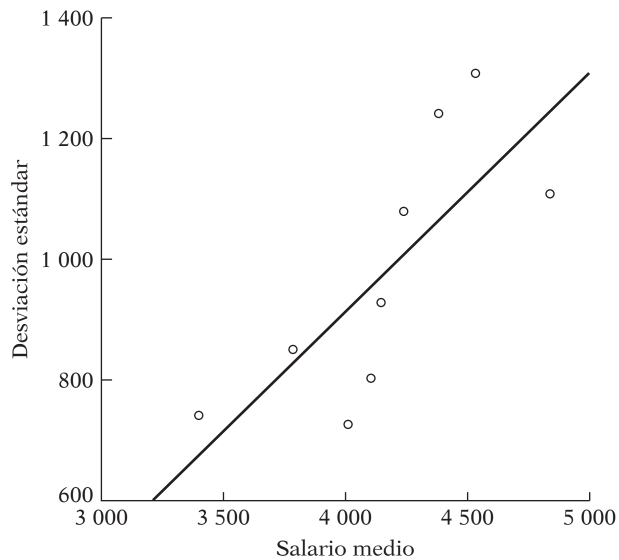
Aunque las industrias difieran en la composición de su producción, la tabla 11.1 muestra con claridad que, en promedio, las empresas grandes pagan más que las pequeñas. Como ejemplo, las empresas que emplean entre uno y cuatro empleados pagaron, en promedio, sueldos de alrededor de \$3 396, mientras que las que emplean entre 1 000 y 2 499 empleados, en promedio, alrededor de \$4 843. Pero observe que hay una gran variabilidad en los ingresos entre las diversas clases de empleados, como indican las desviaciones estándar estimadas de los salarios. Esto también se

**TABLA 11.1** **Salario por empleado (dólares) en la industria de manufactura de bienes no duraderos, de acuerdo con el número de empleados del establecimiento, 1958**

| | Tamaño de la plantilla laboral (número promedio de empleados) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Industria** | **1-4** | **5-9** | **10-19** | **20-49** | **50-99** | **100-249** | **250-499** | **500-999** | **1 000-2 499** |
| Alimentos y similares | 2 994 | 3 295 | 3 565 | 3 907 | 4 189 | 4 486 | 4 676 | 4 968 | 5 342 |
| Productos del tabaco | 1 721 | 2 057 | 3 336 | 3 320 | 2 980 | 2 848 | 3 072 | 2 969 | 3 822 |
| Productos textiles | 3 600 | 3 657 | 3 674 | 3 437 | 3 340 | 3 334 | 3 225 | 3 163 | 3 168 |
| Ropa y productos relacionados | 3 494 | 3 787 | 3 533 | 3 215 | 3 030 | 2 834 | 2 750 | 2 967 | 3 453 |
| Papel y similares | 3 498 | 3 847 | 3 913 | 4 135 | 4 445 | 4 885 | 5 132 | 5 342 | 5 326 |
| Impresión y publicación | 3 611 | 4 206 | 4 695 | 5 083 | 5 301 | 5 269 | 5 182 | 5 395 | 5 552 |
| Productos químicos y similares | 3 875 | 4 660 | 4 930 | 5 005 | 5 114 | 5 248 | 5 630 | 5 870 | 5 876 |
| Productos petroleros y carboníferos | 4 616 | 5 181 | 5 317 | 5 337 | 5 421 | 5 710 | 6 316 | 6 455 | 6 347 |
| Productos de caucho y plásticos | 3 538 | 3 984 | 4 014 | 4 287 | 4 221 | 4 539 | 4 721 | 4 905 | 5 481 |
| Cuero y productos de cuero | 3 016 | 3 196 | 3 149 | 3 317 | 3 414 | 3 254 | 3 177 | 3 346 | 4 067 |
| Remuneración promedio | 3 396 | 3 787 | 4 013 | 4 104 | 4 146 | 4 241 | 4 388 | 4 538 | 4 843 |
| Desviación estándar | 742.2 | 851.4 | 727.8 | 805.06 | 929.9 | 1 080.6 | 1 241.2 | 1 307.7 | 1 110.7 |
| Productividad promedio | 9 355 | 8 584 | 7 962 | 8 275 | 8 389 | 9 418 | 9 795 | 10 281 | 11 750 |

Fuente: *The Census of Manufacturers*, U.S. Department of Commerce, 1958 (calculado por el autor).

**FIGURA 11.6**
Desviación estándar del salario y salario medio.




aprecia en la figura 11.6, que muestra la desviación estándar de los salarios y el salario medio en cada clase de empleados. Como se ve claramente, en promedio, la desviación estándar de los salarios crece con el valor promedio de los salarios.

## 11.2 Estimación por MCO en presencia de heteroscedasticidad

¿Qué sucede con los estimadores de MCO y sus varianzas si introducimos la heteroscedasticidad permitiendo que $E(u_i^2) = \sigma_i^2$ pero conservamos todos los demás supuestos del modelo clásico? Para responder, recuerde el modelo con dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Al aplicar la fórmula usual, el estimador de MCO de $\beta_2$ es

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}\end{aligned} \tag{11.2.1}$$

pero su varianza está dada ahora por la siguiente expresión (apéndice 11A, sección 11A.1):

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{11.2.2}$$

que obviamente difiere de la fórmula usual de varianza obtenida según el supuesto de homoscedasticidad, es decir,

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{11.2.3}$$

Desde luego, si $\sigma_i^2 = \sigma^2$ para cada $i$, las dos fórmulas serán idénticas. (¿Por qué?)

Recuerde que $\hat{\beta}_2$ es el mejor estimador lineal e insesgado (MELI) si se mantienen los supuestos del modelo clásico, inclusive el de homoscedasticidad. ¿Seguirá siendo MELI aunque sólo eliminemos el supuesto de homoscedasticidad y lo reemplacemos por el de heteroscedasticidad? Es fácil probar que $\hat{\beta}_2$ es aún lineal e insesgado. En realidad, como se indica en el apéndice 3A, sección 3A.2, para establecer el insesgamiento de $\hat{\beta}_2$ no es necesario que las perturbaciones ($u_i$) sean homoscedásticas. En realidad, la varianza de $u_i$, homoscedástica o heteroscedástica, no desempeña papel alguno en la determinación de la propiedad de insesgamiento. Recuerde que en el apéndice 3A, sección 3A.7, vimos que $\hat{\beta}_2$ es un estimador consistente según los supuestos del modelo clásico de regresión lineal. Aunque no lo demostraremos, se puede probar que $\hat{\beta}_2$ es un estimador consistente a pesar de la heteroscedasticidad; es decir, conforme el tamaño de la muestra se incrementa de manera indefinida, el $\beta_2$ estimado converge a su valor verdadero. Es más, también puede demostrarse que, en ciertas condiciones (de regulación), $\hat{\beta}_2$ está *distribuida de manera asintótica y normal*. Por supuesto, lo que afirmamos respecto de $\hat{\beta}_2$ también vale para otros parámetros de un modelo de regresión múltiple.

En el supuesto de que $\hat{\beta}_2$ continúe siendo lineal, insesgado y consistente, ¿es "eficiente" o "el mejor"? Es decir, ¿tendrá varianza mínima en la clase de los estimadores lineales e insesgados? ¿Y dicha varianza mínima estará dada por la ecuación (11.2.2)? La respuesta a ambas preguntas es *no*: $\hat{\beta}_2$ deja de ser el mejor y la varianza mínima ya no está dada por (11.2.2). Entonces, ¿cuál estimador es MELI en presencia de heteroscedasticidad? La respuesta se da en la siguiente sección.

## 11.3 El método de mínimos cuadrados generalizados (MCG)

¿Por qué el estimador usual de MCO de $\beta_2$ dado en (11.2.1) no es el mejor, aunque todavía sea insesgado? La razón se intuye en la tabla 11.1. Como se indica ahí, hay una gran variabilidad en los ingresos entre clases de empleados. Si efectuásemos una regresión de salarios por empleado sobre el tamaño de la planta laboral sería recomendable aprovechar el conocimiento de que existe considerable variabilidad entre los salarios de las diferentes clases. Idealmente, querríamos diseñar un esquema de estimación de manera que las observaciones que surgen de poblaciones con mayor variabilidad reciban menos peso que las que provienen de poblaciones con menor variabilidad. Al examinar la tabla 11.1 nos gustaría dar mayor ponderación a las observaciones que provienen de las clases de empleados 10-19 y 20-49 que a las clases de empleados como 5-9 y 250-499, pues las primeras están más concentradas alrededor de sus valores medios que las últimas, lo que permite estimar la FRP en forma más precisa.

Por desgracia, el método de MCO usual no sigue esta estrategia y, por consiguiente, no aprovecha la "información" contenida en la variabilidad desigual de la variable dependiente $Y$, como sucede con la compensación salarial de los empleados de la tabla 11.1: este método asigna igual peso o importancia a cada observación. Pero existe un método de estimación, conocido como **mínimos cuadrados generalizados** (**MCG**), que toma en cuenta esa información explícitamente y, por consiguiente, es capaz de producir estimadores que son MELI. Para ver cómo se hace, considere el modelo ya familiar con dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{11.3.1}$$

el cual, para facilitar el reordenamiento algebraico, escribimos como

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \tag{11.3.2}$$

donde $X_{0i} = 1$ para cada $i$. El lector puede ver que estas dos formulaciones son idénticas.

Ahora, suponga que *se conocen* las varianzas heteroscedásticas $\sigma_i^2$. Divida ambos lados de (11.3.2) entre $\sigma_i$ para obtener

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \tag{11.3.3}$$

la cual, para facilidad de exposición, escribimos como

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \tag{11.3.4}$$

en donde las variables con asterisco o transformadas son las variables originales divididas entre $\sigma_i$ (conocida). Utilizamos la notación $\beta_1^*$ y $\beta_2^*$, los parámetros del modelo transformado, para distinguirlos de los parámetros de MCO usuales $\beta_1$ y $\beta_2$.

¿Cuál es el propósito de transformar el modelo original? Para ver esto, observe la siguiente característica del término de error transformado $u_i^*$:

$$\begin{aligned}
\text{var}\,(u_i^*) = E(u_i^*)^2 &= E\left(\frac{u_i}{\sigma_i}\right)^2 && \text{porque } E(u_i^*) = 0 \\
&= \frac{1}{\sigma_i^2}E\left(u_i^2\right) && \text{porque se conoce } \sigma_i^2 \\
&= \frac{1}{\sigma_i^2}(\sigma_i^2) && \text{porque } E\left(u_i^2\right) = \sigma_i^2 \\
&= 1
\end{aligned} \tag{11.3.5}$$

que es una constante. Es decir, la varianza del término de perturbación transformado $u_i^*$ es ahora homoscedástica. Como aún se conservan los demás supuestos del modelo clásico, el hallazgo de que $u^*$ es homoscedástico sugiere que si se aplican MCO al modelo transformado (11.3.3) se producirán estimadores MELI. En resumen, los $\beta_1^*$ y $\beta_2^*$ estimados son ahora MELI y no los estimadores de MCO, $\hat{\beta}_1$ y $\hat{\beta}_2$.

Este procedimiento de transformar las variables originales de forma que las variables transformadas satisfagan los supuestos del modelo clásico y de aplicar luego MCO a ellos se conoce como método de mínimos cuadrados generalizados (MCG). *En resumen, MCG es MCO sobre las variables transformadas que satisfacen los supuestos estándar de mínimos cuadrados.* Los estimadores así obtenidos se conocen como **estimadores de MCG**, y son estos estimadores los que son MELI.

El mecanismo de estimación de $\beta_1^*$ y $\beta_2^*$ es el siguiente. Primero escribimos la FRM de (11.3.3)

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{\hat{u}_i}{\sigma_i}\right)$$

o

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{11.3.6}$$

Ahora, para obtener los estimadores de MCG, reducimos

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

es decir,

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right)\right]^2 \tag{11.3.7}$$

El mecanismo real para minimizar (11.3.7) sigue las técnicas estándar del cálculo y se presenta en el apéndice 11A, sección 11A.2. Como allí se muestra, el estimador de MCG de $\beta_2^*$ es

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \tag{11.3.8}$$

y su varianza está dada por

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \tag{11.3.9}$$

donde $w_i = 1/\sigma_i^2$.

## Diferencia entre MCO y MCG

Recuerde, del capítulo 3, que en MCO reducimos

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{11.3.10}$$

pero en MCG reducimos la expresión (11.3.7), la cual también se escribe como

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2 \tag{11.3.11}$$

donde $w_i = 1/\sigma_i^2$ [verifique que (11.3.11) y (11.3.7) sean idénticas].

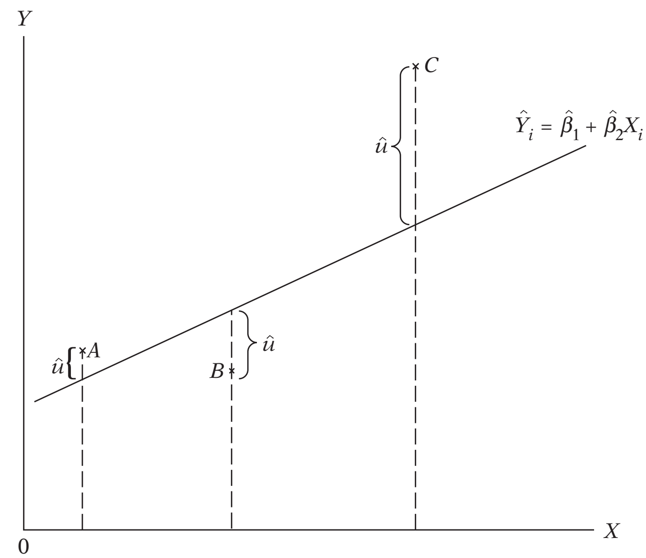
Por tanto, en MCG se reduce una *suma ponderada de residuos al cuadrado* donde $w_i = 1/\sigma_i^2$ actúan como ponderación, pero en MCO se reduce la SCR sin ponderar o (lo que equivale a lo mismo) con ponderaciones iguales. Como muestra (11.3.7), en MCG, el peso asignado a cada observación es inversamente proporcional a su $\sigma_i$, es decir, las observaciones que provienen de una población con una $\sigma_i$ más grande tendrán una ponderación relativamente menor, y las de una población con un $\sigma_i$ menor tendrán una ponderación proporcionalmente mayor al reducir la SCR (11.3.11). Para ver claramente la diferencia entre MCO y MCG considere el diagrama hipotético de dispersión de la figura 11.7.

En MCO (sin ponderar), cada $\hat{u}_i^2$ asociada con los puntos $A$, $B$ y $C$ recibirá el mismo peso al reducir la SCR. Obviamente, en este caso la $\hat{u}_i^2$ asociada al punto $C$ dominará la SCR. Pero en MCG la observación extrema $C$ obtendrá relativamente un peso menor que las otras dos observaciones. Como ya mencionamos, ésta es la estrategia correcta, pues para estimar la función de regresión poblacional (FRP) de una manera más confiable, sería deseable dar más peso a las observaciones agrupadas cerca de su media (poblacional) que a las que están muy dispersas a su alrededor.

Como (11.3.11) reduce una SCR ponderada, esto se conoce apropiadamente como **mínimos cuadrados ponderados** (**MCP**), y los estimadores así obtenidos que aparecen en (11.3.8) y (11.3.9) se conocen como **estimadores de MCP**. Pero MCP es apenas un caso especial de la técnica de estimación más general, MCG. En el contexto de la heteroscedasticidad los términos MCP y MCG se tratan indistintamente. En capítulos posteriores veremos otros casos especiales de MCG.

A propósito, observe que si $w_i = w$, una constante para todas las $i$, $\hat{\beta}_2^*$ es idéntico a $\hat{\beta}_2$ y var ($\hat{\beta}_2^*$) es idéntica a la var ($\hat{\beta}_2$) usual (es decir, homoscedástica) dada en (11.2.3), lo cual no debe sorprender. (¿Por qué?) (Consulte el ejercicio 11.8.)

**FIGURA 11.7**
Diagrama de dispersión hipotético.




## 11.4 Consecuencias de utilizar MCO en presencia de heteroscedasticidad

Como vimos, $\hat{\beta}_2^*$ y $\hat{\beta}_2$ son estimadores (lineales) insesgados: para muestreo repetido, en promedio, $\hat{\beta}_2^*$ y $\hat{\beta}_2$ serán iguales al verdadero $\beta_2$, es decir, ambos son estimadores insesgados. Pero sabemos que $\hat{\beta}_2^*$ es el eficiente, es decir, tiene la menor varianza. ¿Qué sucede con los intervalos de confianza, las pruebas de hipótesis y con otros procedimientos si continuamos utilizando el estimador de MCO, $\hat{\beta}_2$? Se distinguen dos situaciones.

### Estimación por MCO con heteroscedasticidad

Suponga que utilizamos $\hat{\beta}_2$ y la fórmula de varianza dada en (11.2.2), la cual considera explícitamente la heteroscedasticidad. Con esta varianza y la suposición de que se conocen las $\sigma_i^2$, ¿es posible establecer intervalos de confianza y probar hipótesis con las pruebas $t$ y $F$ usuales? La respuesta suele ser no, pues puede demostrarse que var $(\hat{\beta}_2^*) \leq$ var $(\hat{\beta}_2)$,[5] lo cual significa que los intervalos de confianza basados en estos últimos serán innecesariamente grandes. Como resultado, es probable que las pruebas $t$ y $F$ den resultados imprecisos en el sentido de que la var $(\hat{\beta}_2)$ es demasiado grande, y lo que parece un coeficiente estadísticamente no significativo (pues el valor $t$ es más bajo de lo apropiado), de hecho puede resultar significativo si se establecen intervalos de confianza correctos con base en el procedimiento de MCG.

### Estimación por MCO sin heteroscedasticidad

La situación se torna muy grave si, además de $\hat{\beta}_2$, también se sigue utilizando la fórmula habitual de varianza (homoscedástica) dada en (11.2.3), aunque exista heteroscedasticidad o se sospeche su existencia: observe que éste es el caso más probable de los dos que aquí se analizan, pues al hacer una regresión estándar por MCO e ignorar (o no conocer) la existencia de la heteroscedasticidad se producirá una varianza de $\hat{\beta}_2$ como la dada en (11.2.3). En primer lugar, la var $(\hat{\beta}_2)$ dada en (11.2.3) es un estimador *sesgado* de var $(\hat{\beta}_2)$ dada en (11.2.2), es decir, en promedio, ésta

[5] Una prueba formal se encuentra en Phoebus J. Dhrymes, *Introductory Econometrics,* Springer-Verlag, Nueva York, 1978, pp. 110-111. A propósito, observe que la pérdida de eficiencia de $\hat{\beta}_2$ [es decir, la medida en que excede var $[\hat{\beta}_2]$ a var $[\hat{\beta}_2^*]$ depende de los valores muestrales de las variables $X$ y del valor de $\sigma_i^2$.

sobreestima o subestima la última y *en general* no se puede decir si el sesgo es positivo (sobreestimación) o negativo (subestimación), pues éste depende de la naturaleza de la relación entre $\sigma_i^2$ y los valores tomados por la variable explicativa $X$, como se ve claramente en (11.2.2) (véase el ejercicio 11.9). El sesgo surge de que $\hat{\sigma}^2$, el estimador convencional de $\sigma^2$, a saber, $\sum \hat{u}_i^2/(n-2)$, deja de ser un estimador insesgado del último en presencia de heteroscedasticidad (véase el apéndice 11A.3). Como resultado, ya no es posible depender de los intervalos de confianza calculados convencionalmente ni de las pruebas $t$ y $F$ tradicionales.[6] **En resumen, si insistimos en los procedimientos de prueba usuales a pesar de la presencia de heteroscedasticidad, las conclusiones o inferencias que obtengamos pueden ser muy equivocadas**.

Para dar mayor claridad a este tema, veamos un estudio **Monte Carlo** de Davidson y MacKinnon.[7] Consideran el siguiente modelo simple, que en nuestra notación es

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{11.4.1}$$

Suponen que $\beta_1 = 1$, $\beta_2 = 1$ y $u_i \sim N(0, X_i^\alpha)$. Como indica la última expresión, suponen que la varianza del error es heteroscedástica y que se relaciona con el valor de la regresora $X$ elevado a la potencia $\alpha$. Si, por ejemplo, $\alpha = 1$, la varianza del error es proporcional al valor de $X$; si $\alpha = 2$, la varianza del error es proporcional al cuadrado del valor de $X$, y así sucesivamente. En la sección 11.6 veremos la lógica de tal procedimiento. Basados en 20 000 repeticiones y con diversos valores para $\alpha$, obtienen los errores estándar de los dos coeficientes de regresión con MCO [consulte la ecuación (11.2.3)], MCO con heteroscedasticidad [ecuación (11.2.2)] y MCG [ecuación (11.3.9)]. Sus resultados para valores seleccionados de $\alpha$ son:

| | Error estándar de $\hat{\beta}_1$ | | | Error estándar de $\hat{\beta}_2$ | | |
|---|---|---|---|---|---|---|
| **Valor de $\alpha$** | **MCO** | **$MCO_{het}$** | **MCG** | **MCO** | **$MCO_{het}$** | **MCG** |
| 0.5 | 0.164 | 0.134 | 0.110 | 0.285 | 0.277 | 0.243 |
| 1.0 | 0.142 | 0.101 | 0.048 | 0.246 | 0.247 | 0.173 |
| 2.0 | 0.116 | 0.074 | 0.0073 | 0.200 | 0.220 | 0.109 |
| 3.0 | 0.100 | 0.064 | 0.0013 | 0.173 | 0.206 | 0.056 |
| 4.0 | 0.089 | 0.059 | 0.0003 | 0.154 | 0.195 | 0.017 |

*Nota:* $MCO_{het}$ significa MCO con heteroscedasticidad.

*La característica más sobresaliente de estos resultados es que los MCO, con o sin corrección por heteroscedasticidad, sobreestiman consistentemente el verdadero error estándar obtenido mediante el procedimiento (correcto) de MCG, en especial para valores grandes de α, con lo cual se establece la superioridad de MCG.* Estos resultados también muestran que si no se utilizan MCG y se depende de MCO —permitiendo o no la heteroscedasticidad—, el resultado es una mezcla. Los errores estándar de MCO convencionales son muy grandes (para el intercepto) o *generalmente* muy bajos (para el coeficiente de la pendiente) en relación con los obtenidos por MCO que permite la heteroscedasticidad. El mensaje es claro: ante la presencia de heteroscedasticidad, utilice MCG. Sin embargo, por razones explicadas más adelante en el capítulo, en la práctica no siempre es fácil aplicar MCG. Asimismo, como veremos más adelante, a menos que la heteroscedasticidad sea muy grave, no se abandonarían los MCO en favor de los MCG o de los MCP.

Por el análisis anterior, es claro que la heteroscedasticidad es un problema potencialmente grave y el investigador debe saber si está presente en una situación dada. Si detecta su presen-

[6] De (5.3.6) sabemos que el intervalo de confianza a $100(1-\alpha)\%$ para $\beta_2$ es $[\hat{\beta}_2 \pm t_{\alpha/2}\,\text{ee}(\hat{\beta}_2)]$. Pero, si $\text{ee}(\hat{\beta}_2)$ no puede estimarse en forma insesgada, ¿qué confianza se puede tener en el intervalo de confianza calculado convencionalmente?

[7] Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics,* Oxford University Press, Nueva York, 1993, pp. 549-550.

cia, puede tomar acciones correctivas, como una regresión de mínimos cuadrados ponderados o alguna otra técnica. Sin embargo, antes de examinar los procedimientos correctivos, es preciso averiguar si hay heteroscedasticidad o si es probable en un caso dado. Analizamos este tema en la siguiente sección.

### Nota técnica

Aunque ya establecimos que, en caso de heteroscedasticidad, son los MCG y no los MCO los que son MELI, existen ejemplos en los que los MCO pueden ser MELI a pesar de la heteroscedasticidad.[8] No obstante, dichos casos son poco frecuentes en la práctica.

## 11.5 Detección de la heteroscedasticidad

Como con la multicolinealidad, para efectos prácticos la pregunta importante es ¿cómo saber que hay heteroscedasticidad en una situación específica? De nuevo, como en la multicolinealidad, no existen reglas precisas y rápidas para detectar la heteroscedasticidad, sólo algunas reglas prácticas. Pero esta situación es inevitable porque $\sigma_i^2$ solamente puede conocerse si se tiene toda la población $Y$, correspondiente a las $X$ seleccionadas, como la población presentada en la tabla 2.1 o en la tabla 11.1. Pero tal información es una excepción más que la regla en la mayoría de las investigaciones económicas. Al respecto, el econometrista difiere de los científicos en campos como agricultura y biología, donde los investigadores tienen gran parte del control sobre sus temas. En los estudios de economía, es frecuente que sólo haya un valor muestral $Y$ correspondiente a un valor particular de $X$. Por consiguiente, no hay forma de conocer $\sigma_i^2$ a partir de una sola observación $Y$. Así, en la mayoría de las investigaciones econométricas, la heteroscedasticidad puede ser un asunto de intuición, de conjeturas refinadas, de un trabajo basado en experiencia empírica previa o de pura especulación.

Tras la advertencia anterior, podemos examinar algunos métodos informales y formales para detectar la heteroscedasticidad. Como revelará el siguiente análisis, la mayoría de estos métodos se basan en el examen de los residuos $\hat{u}_i$ de MCO, pues son éstos los que se observan y no las perturbaciones $u_i$. Se espera que $\hat{u}_i$ sean buenas estimaciones de $u_i$, esperanza que se cumple si el tamaño de la muestra es lo bastante grande.

### Métodos informales

*Naturaleza del problema*

Con mucha frecuencia la naturaleza del problema en consideración sugiere la posibilidad de heteroscedasticidad. Por ejemplo, a partir del trabajo pionero de Prais y Houthakker sobre estudios de presupuesto familiar, en el cual hallaron que la varianza residual correspondiente a la regresión del consumo sobre el ingreso aumentaba con el ingreso, hoy en día generalmente se supone que en encuestas similares se pueden esperar varianzas desiguales entre las perturbaciones.[9] De hecho, en la información de corte transversal que comprende unidades heterogéneas, la heteroscedasticidad puede ser la regla y no la excepción. Así, en el análisis de corte transversal que relaciona el gasto de inversión con las ventas, la tasa de interés, etc., suele esperarse la presencia de heteroscedasticidad si se agrupan empresas pequeñas, medianas y grandes.

[8] La razón reside en que el teorema de Gauss-Markov proporciona la condición suficiente (pero no necesaria) para que los MCO sean eficientes. La condición suficiente y necesaria para que los MCO sean MELI la establece el **teorema de Kruskal**. Pero este tema está más allá del alcance de este libro. El autor está en deuda con Michael McAleer por notar esto. Para mayores detalles, véase Denzil G. Fiebig, Michael McAleer y Robert Bartels, "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances", *Journal of Econometrics,* vol. 54, núm. 1-3, octubre-diciembre de 1992, pp. 321-334. Para los estudiantes con inclinaciones matemáticas, analizo este tema en más detalle en el **apéndice C** con álgebra matricial.

[9] S.J. Prais y H.S. Houthakker, *The Analysis of Family Budgets,* Cambridge University Press, Nueva York, 1955.

De hecho, ya presentamos ejemplos de lo anterior. En el capítulo 2 analizamos la relación entre los salarios medios o promedios por hora en relación con los años de escolaridad en Estados Unidos. En este capítulo también examinamos la relación entre el gasto en alimentos y el gasto total de 55 familias en India (ejercicio 11.16).

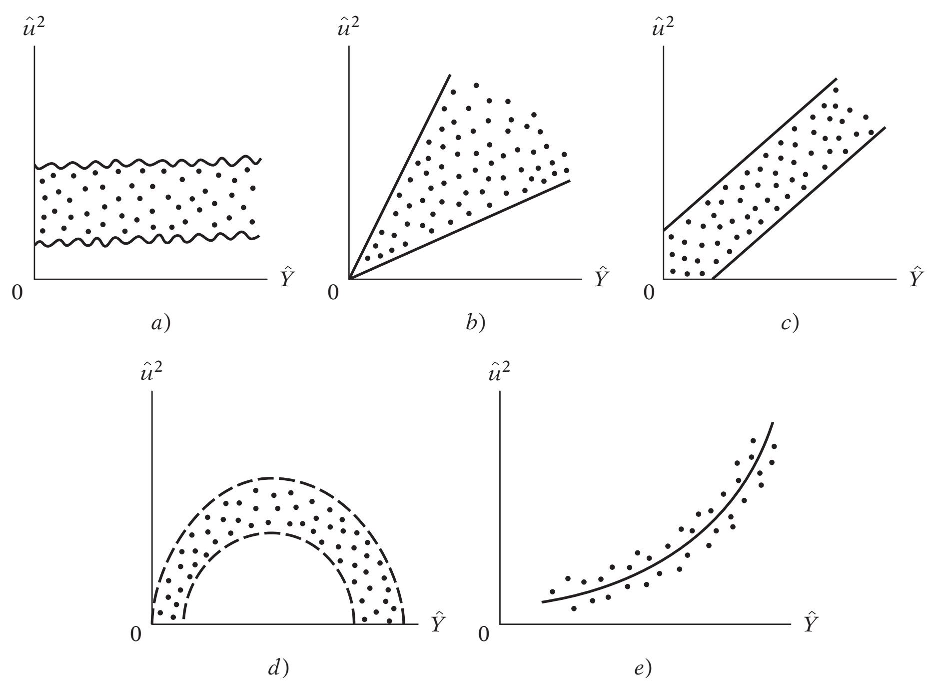
### *Método gráfico*

Si no hay información *a priori* o empírica sobre la naturaleza de la heteroscedasticidad, en la práctica se puede llevar a cabo un análisis de regresión con el supuesto de que no hay heteroscedasticidad y luego hacer un examen *post mortem* de los residuos elevados al cuadrado, $\hat{u}_i^2$, para ver si exhiben algún patrón sistemático. Aunque los $\hat{u}_i^2$ no son lo mismo que los $u_i^2$, los primeros sirven como representantes de los últimos sobre todo si el tamaño de la muestra es lo bastante grande.[10] Un examen de los $\hat{u}_i^2$ puede revelar patrones como los de la figura 11.8.

En la figura 11.8 se grafican los $\hat{u}_i^2$ frente a los $\hat{Y}_i$, que son los $Y_i$ estimados mediante la línea de regresión, con la idea de averiguar si el valor medio estimado de $Y$ está relacionado sistemáticamente con el residuo al cuadrado. En la figura 11.8*a*) se ve que no hay un patrón sistemático entre las dos variables, lo cual sugiere que tal vez no haya heteroscedasticidad en los datos. Sin embargo, las figuras 11.8*b*) a 11.8*e*) muestran patrones definidos. Por ejemplo, la figura 11.8*c*) sugiere una relación lineal, mientras que las figuras 11.8*d*) y 11.8*e*) indican una relación cuadrática entre $\hat{u}_i^2$ y $\hat{Y}_i$. Con tal conocimiento, aunque informal, es posible transformar los datos de manera que, una vez transformados, no presenten heteroscedasticidad. En la sección 11.6 examinaremos transformaciones de este tipo.

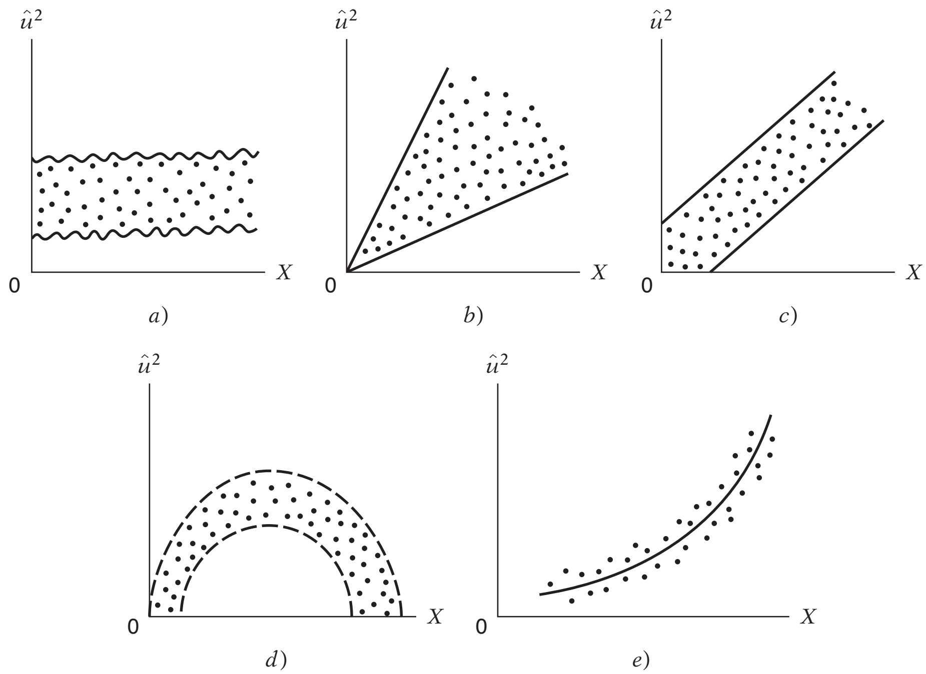
En lugar de graficar los $\hat{u}_i^2$ frente a los $\hat{Y}_i$, se pueden graficar frente a una de las variables explicativas, en especial si el gráfico de $\hat{u}_i^2$ frente a $\hat{Y}_i$ presenta un patrón como el de la figura 11.8*a*). Tal gráfico, que aparece en la figura 11.9, puede revelar patrones similares a los de la figura 11.8. (En el caso del modelo con dos variables, el gráfico de $\hat{u}_i^2$ frente a los $\hat{Y}_i$ equivale a graficar el

**FIGURA 11.8**
Patrones hipotéticos de los residuos estimados al cuadrado.




[10] Respecto de la relación entre $\hat{u}_i$ y $u_i$, véase E. Malinvaud, *Statistical Methods of Econometrics,* North Holland, Amsterdam, 1970, pp. 88-89.

**FIGURA 11.9**
Diagrama de dispersión de los residuos estimados al cuadrado frente a $X$.




primero frente a $X_i$, razón por la cual la figura 11.9 es similar a la 11.8. Pero ésta no es la situación cuando se considera un modelo con dos o más variables $X$; en este caso, $\hat{u}_i^2$ puede graficarse frente a cualquier variable $X$ incluida en el modelo.)

Un patrón como el de la figura 11.9*c*), por ejemplo, sugiere que la varianza del término de perturbación se relaciona linealmente con la variable $X$. Así, si en la regresión de ahorro sobre ingreso se encuentra un patrón como el de la figura 11.9*c*), éste sugiere que la varianza heteroscedástica puede ser *proporcional* al valor de la variable ingreso. Saber esto puede contribuir a transformar los datos de manera que, en la regresión sobre los datos transformados, la varianza de las perturbaciones sea homoscedástica. Trataremos este tema en la siguiente sección.

## Métodos formales

### *Prueba de Park*[11]

Park formaliza el método gráfico con la sugerencia de que $\sigma_i^2$ es algún tipo de función de la variable explicativa $X_i$. La forma funcional fue

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{vi}$$

o

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \tag{11.5.1}$$

donde $v_i$ es el término de perturbación estocástico.

[11] R.E. Park, "Estimation with Heteroscedastic Error Terms", *Econometrica*, vol. 34, núm. 4, octubre de 1966, p. 888. La prueba de Park es un caso especial de la prueba general propuesta por A. C. Harvey en "Estimating Regression Models with Multiplicative Heteroscedasticity", *Econometrica,* vol. 44, núm. 3, 1976, pp. 461-465.

Como $\sigma_i^2$ por lo general no se conoce, Park sugiere utilizar $\hat{u}_i^2$ como aproximación y correr la siguiente regresión:

$$\begin{aligned}\ln \hat{u}_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\ &= \alpha + \beta \ln X_i + v_i\end{aligned} \tag{11.5.2}$$

Si $\beta$ resulta estadísticamente significativo, esto sugerirá heteroscedasticidad en los datos. Si resulta no significativo, podemos aceptar el supuesto de homoscedasticidad. La prueba de Park es, por tanto, un procedimiento de dos etapas. En la primera se efectúa la regresión MCO ignorando el interrogante de la heteroscedasticidad. Se obtiene $\hat{u}_i$ de esta regresión y luego, en la segunda etapa, se efectúa la regresión (11.5.2).

Aunque empíricamente la prueba de Park es atractiva, presenta algunos problemas. Goldfeld y Quandt argumentan que el término de error $v_i$ que entra en (11.5.2) puede no satisfacer los supuestos de MCO y en sí mismo ser heteroscedástico.[12] No obstante, es posible utilizar la prueba de Park como método estrictamente exploratorio.

**EJEMPLO 11.1**
*Relación entre salarios y productividad*

Para ilustrar el enfoque de Park, con la información de la tabla 11.1 efectuamos la siguiente regresión:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

donde $Y$ = salario promedio en miles de dólares, $X$ = productividad promedio en miles de dólares e $i$ = $i$-ésimo de la planta laboral del establecimiento. Los resultados de la regresión fueron los siguientes:

$$\begin{aligned}\hat{Y}_i &= 1992.3452 + 0.2329X_i \\ \text{ee} &= (936.4791) \quad (0.0998) \\ t &= \quad (2.1275) \quad (2.333) \qquad R^2 = 0.4375\end{aligned} \tag{11.5.3}$$

Los resultados revelan que el coeficiente de pendiente estimado es significativo en el nivel de 5% con base en una prueba $t$ de una cola. La ecuación muestra que, a medida que aumenta la productividad laboral, por ejemplo, en un dólar, el salario aumenta, en promedio, alrededor de 23 centavos de dólar.

En los residuos obtenidos de la regresión (11.5.3) se hizo la regresión sobre $X_i$ como lo sugiere la ecuación (11.5.2), con los siguientes resultados:

$$\begin{aligned}\widehat{\ln \hat{u}_i^2} &= 35.817 - 2.8099 \ln X_i \\ \text{ee} &= (38.319) \quad (4.216) \\ t &= (0.934) \quad (-0.667) \qquad R^2 = 0.0595\end{aligned} \tag{11.5.4}$$

Obvio, no hay una relación estadísticamente significativa entre ambas variables. Según la prueba de Park, se puede concluir que no hay heteroscedasticidad en la varianza del error.[13]

### *Prueba de Glejser*[14]

La prueba de Glejser en esencia es similar a la de Park. Después de obtener los residuos $\hat{u}_i$ de la regresión MCO, Glejser sugiere una regresión sobre los valores absolutos de $\hat{u}_i$ sobre la variable

[12] Stephen M. Goldfeld y Richard E. Quandt, *Nonlinear Methods in Econometrics,* North Holland, Amsterdam, 1972, pp. 93-94.

[13] La forma funcional particular seleccionada por Park es sólo una sugerencia. Una forma funcional diferente revela relaciones significativas. Por ejemplo, se puede utilizar $\hat{u}_i^2$ en lugar de ln $\hat{u}_i^2$ como variable dependiente.

[14] H. Glejser, "A New Test for Heteroscedasticity", *Journal of the American Statistical Association,* vol. 64, 1969, pp. 316-323.

$X$ que se cree muy asociada con $\sigma_i^2$. En sus experimentos, Glejser utilizó las siguientes formas funcionales:

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

donde $v_i$ es el término de error.

De nuevo, como un asunto empírico o práctico, se puede utilizar el método de Glejser. Sin embargo, Goldfeld y Quandt señalan que el término de error $v_i$ tiene algunos problemas, pues su valor esperado es diferente de cero, está serialmente correlacionado (véase el capítulo 12) e irónicamente es heteroscedástico.[15] Otra dificultad del método Glejser es que los modelos como

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

y

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

no son lineales en los parámetros y, por consiguiente, no pueden estimarse mediante el procedimiento de MCO habitual.

Glejser descubrió que para muestras grandes, los cuatro primeros modelos suelen dar resultados satisfactorios en la detección de la heteroscedasticidad. En la práctica, por consiguiente, la técnica de Glejser es útil para muestras grandes, y en muestras pequeñas sirve estrictamente como herramienta cualitativa para obtener una noción sobre la heteroscedasticidad.

**EJEMPLO 11.2**
*Relación entre el salario y la productividad: prueba de Glejser*

Aún con el ejemplo 11.1, se hizo la regresión del valor absoluto de los residuos obtenidos de la regresión (11.5.3) sobre la productividad promedio ($X$), lo cual dio los siguientes resultados:

$$\begin{aligned} \widehat{|\hat{u}_i|} &= 407.2783 \quad - \quad 0.0203X_i \\ \text{ee} &= (633.1621) \quad (0.0675) \qquad r^2 = 0.0127\text{e} \\ t &= (0.6432) \quad (-0.3012) \end{aligned} \tag{11.5.5}$$

Como se aprecia de esta regresión, no hay relación entre el valor absoluto de los residuos y la regresora, la productividad promedio. Lo anterior refuerza la conclusión basada en la prueba de Park.

### *Prueba de correlación de orden de Spearman*

En el ejercicio 3.8 definimos el coeficiente de correlación de orden de Spearman como

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)}\right] \tag{11.5.6}$$

[15] Véanse detalles en Goldfeld y Quandt, *op. cit.*, capítulo 3.

donde $d_i$ = la diferencia en las posiciones o lugares asignados al $i$-ésimo individuo o fenómeno respecto de dos características y $n$ = número de individuos o fenómenos ordenados. Con el coeficiente de correlación de orden anterior se detecta heteroscedasticidad de la siguiente manera: Suponga que $Y_i = \beta_0 + \beta_1 X_i + u_i$.

**Paso 1.** Ajuste la regresión a los datos sobre $Y$ y $X$, y obtenga los residuos $\hat{u}_i$.

**Paso 2.** Ignore el signo de $\hat{u}_i$, es decir, tome su valor absoluto $|\hat{u}_i|$, y ordene los valores $|\hat{u}_i|$ y $X_i$ (o $\hat{Y}_i$) de acuerdo con un orden ascendente o descendente, y calcule el coeficiente de correlación de orden de Spearman dado antes.

**Paso 3.** Si supone que el coeficiente poblacional de correlación de orden $\rho_s$ es cero y $n > 8$, la significancia del $r_s$ muestral se prueba mediante la prueba $t$ de la siguiente manera:[16]

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \tag{11.5.7}$$

con gl $= n - 2$.

Si el valor $t$ calculado excede el valor $t$ crítico, podemos aceptar la hipótesis de heteroscedasticidad; de lo contrario, podemos rechazarla. Si el modelo de regresión considera más de una variable $X$, $r_s$ se calcula entre $|\hat{u}_i|$ y cada variable $X$ por separado, y la significancia estadística se somete a la prueba $t$ dada en la ecuación (11.5.7).

**EJEMPLO 11.3**
*Ilustración de la prueba de correlación de orden*

Para ilustrar la prueba de correlación de orden, considere los datos de la tabla 11.2, que corresponden al rendimiento anual promedio ($E$, %) y la desviación estándar del rendimiento anual ($\sigma_i$,%) de 10 fondos de inversión.

**TABLA 11.2** **Prueba de correlación de orden para heteroscedasticidad**

| Nombre del fondo mutualista | $E_i$, rendimiento promedio anual, % | $\sigma_i$, desviación estándar del rendimiento anual, % | $\hat{E}_i$† | $\lvert\hat{u}_i\rvert$‡, residuos $\lvert(E_i - \hat{E}_i)\rvert$ | Ordenación de $\lvert\hat{u}_i\rvert$ | Ordenación de $\sigma_i$ | $d$, diferencia entre las dos ordenaciones | $d^2$ |
|---|---|---|---|---|---|---|---|---|
| Boston Fund | 12.4 | 12.1 | 11.37 | 1.03 | 9 | 4 | 5 | 25 |
| Delaware Fund | 14.4 | 21.4 | 15.64 | 1.24 | 10 | 9 | 1 | 1 |
| Equity Fund | 14.6 | 18.7 | 14.40 | 0.20 | 4 | 7 | −3 | 9 |
| Fundamental Investors | 16.0 | 21.7 | 15.78 | 0.22 | 5 | 10 | −5 | 25 |
| Investors Mutual | 11.3 | 12.5 | 11.56 | 0.26 | 6 | 5 | 1 | 1 |
| Loomis-Sales Mutual Fund | 10.0 | 10.4 | 10.59 | 0.59 | 7 | 2 | 5 | 25 |
| Massachusetts Investors Trust | 16.2 | 20.8 | 15.37 | 0.83 | 8 | 8 | 0 | 0 |
| New England Fund | 10.4 | 10.2 | 10.50 | 0.10 | 3 | 1 | 2 | 4 |
| Putnam Fund of Boston | 13.1 | 16.0 | 13.16 | 0.06 | 2 | 6 | −4 | 16 |
| Wellington Fund | 11.3 | 12.0 | 11.33 | 0.03 | 1 | 3 | −2 | 4 |
| Total | | | | | | | 0 | 110 |

† Obtenido de la regresión: $\hat{E}i = 5.8194 + 0.4590\ \sigma_i$.
‡ Valor absoluto de los residuales.
*Nota:* La ordenación de valores es ascendente.

(*continúa*)

[16] Véase G. Udny Yule y M.G. Kendall, *An Introduction to the Theory of Statistics,* Charles Griffin & Company, Londres, 1953, p. 455.

**EJEMPLO 11.3** (*continuación*)

La línea del mercado de capitales (LMC) de la teoría de portafolios postula una relación lineal entre el rendimiento esperado ($E_i$) y el riesgo (como se mide mediante la desviación estándar, $\sigma$) de un portafolio, de la siguiente manera:

$$E_i = \beta_i + \beta_2 \sigma_i$$

Con los datos de la tabla 11.2 se estimó el modelo anterior, a partir del cual se calcularon los residuos. Como los datos se relacionan con 10 fondos mutualistas de distintos tamaños y metas de inversión, se podría esperar *a priori* que hubiera heteroscedasticidad. Para probar esta hipótesis, aplicamos la prueba de correlación de orden. Los cálculos necesarios se proporcionan en la tabla 11.2.

Con la fórmula (11.5.6) obtenemos

$$\begin{aligned} r_s &= 1 - 6\frac{110}{10(100-1)} \\ &= 0.3333 \end{aligned} \tag{11.5.8}$$

Con la prueba $t$ dada en (11.5.7) obtenemos

$$\begin{aligned} t &= \frac{(0.3333)(\sqrt{8})}{\sqrt{1-0.1110}} \\ &= 0.9998 \end{aligned} \tag{11.5.9}$$

Para 8 gl, este valor $t$ no es significativo ni siquiera en el nivel de significancia de 10%; el valor $p$ es 0.17. Por tanto, no hay evidencia de una relación sistemática entre la variable explicativa y los valores absolutos de los residuos, lo cual sugeriría que no hay heteroscedasticidad.

### *Prueba de Goldfeld-Quandt*[17]

Este popular método es aplicable si se supone que la varianza heteroscedástica, $\sigma_i^2$, está relacionada positivamente con *una* de las variables explicativas en el modelo de regresión. Por simplicidad, considere el modelo usual con dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Suponga que $\sigma_i^2$ está relacionado positivamente con $X_i$, en la forma

$$\sigma_i^2 = \sigma^2 X_i^2 \tag{11.5.10}$$

donde $\sigma^2$ es una constante.[18]

El supuesto (11.5.10) postula que $\sigma_i^2$ es proporcional al cuadrado de la variable $X$. En su estudio de presupuestos familiares, Prais y Houthakker encontraron muy útil ese supuesto. (Véase la sección 11.5, métodos informales.)

Si (11.5.10) es la relación apropiada, significaría que $\sigma_i^2$ sería mayor mientras mayores fueran los valores de $X_i$. Si éste resulta ser el caso, es muy probable que haya heteroscedasticidad en el modelo. Para probar esto explícitamente, Goldfeld y Quandt sugieren los siguientes pasos:

**Paso 1.** Ordene las observaciones de acuerdo con los valores de $X_i$, a partir del valor más bajo de $X$.

**Paso 2.** Omita las $c$ observaciones centrales, donde $c$ se especificó *a priori,* y divida las observaciones restantes $(n - c)$ en dos grupos, cada uno de $(n - c)/2$ observaciones.

**Paso 3.** Ajuste regresiones MCO separadas a las primeras $(n - c)/2$ observaciones y a las últimas $(n - c)/2$ observaciones, y obtenga las respectivas sumas de cuadrados residuales

[17] Goldfeld y Quandt, *op. cit.*, capítulo 3.

[18] Éste es sólo un supuesto plausible. En realidad lo que se requiere es que $\sigma_i^2$ esté monotónicamente relacionada con $X_i$.

$SCR_1$ y $SCR_2$; $SCR_1$ representa la SCR de la regresión correspondiente a los valores más bajos de $X_i$ (el grupo de varianza pequeña), y $SCR_2$, a los valores más grandes de $X_i$ (el grupo de varianza grande). Cada SCR tiene

$$\frac{(n-c)}{2}-k \qquad \text{o} \qquad \left(\frac{n-c-2k}{2}\right) \text{ gl}$$

donde $k$ es el número de parámetros que deben estimarse, inclusive el intercepto. (¿Por qué?) Sin duda, para el caso de dos variables, $k$ es 2.

**Paso 4.** Calcule la razón

$$\lambda=\frac{SCR_2/\text{gl}}{SCR_1/\text{gl}} \tag{11.5.11}$$

*Si supusimos que las $u_i$ están normalmente distribuidas* (lo cual suele hacerse), y *si el supuesto de homoscedasticidad es válido,* entonces se demuestra que $\lambda$ de (11.5.10) sigue la distribución $F$ con un número de gl en el numerador y uno en el denominador iguales a $(n-c-2k)/2$.

Si en una aplicación $\lambda\ (=F)$ calculada es superior al $F$ crítico en el nivel de significancia seleccionado, podemos rechazar la hipótesis de homoscedasticidad, es decir, podemos afirmar que la heteroscedasticidad es muy probable.

Antes de ilustrar la prueba, conviene explicar la omisión de las observaciones centrales $c$. Estas observaciones se omiten para agudizar o acentuar la diferencia entre el grupo de varianza pequeña (es decir, $SCR_1$) y el grupo de varianza grande (es decir, $SCR_2$). Pero la capacidad de la prueba Goldfeld-Quandt para lograrlo depende de la forma de seleccionar $c$.[19] Para el modelo con dos variables, los experimentos Monte Carlo realizados por Goldfeld y Quandt sugieren que $c$ sea alrededor de 8 si el tamaño de la muestra es alrededor de 30, y alrededor de 16 si el tamaño de la muestra es alrededor de 60. Sin embargo, Judge *et al*., encontraron satisfactorios en la práctica los niveles de $c=4$ si $n=30$ y $c=10$ si $n$ es alrededor de 60.[20]

Antes de proseguir, cabe notar que, en caso de que haya más de una variable $X$ en el modelo, el ordenamiento de las observaciones, que es el primer paso en la prueba, puede hacerse de acuerdo con cualquiera de ellas. Por tanto, en el modelo: $Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\beta_4X_{4i}+u_i$ se pueden ordenar los datos de acuerdo con cualquiera de estas $X$. Si, *a priori*, no hay seguridad sobre cuál variable $X$ es la adecuada, realice la prueba sobre cada variable $X$ o aplique la prueba de Park, por turnos, sobre cada $X$.

**EJEMPLO 11.4**
*La prueba de Goldfeld-Quandt*

Para ilustrar la prueba de Goldfeld-Quandt presentamos en la tabla 11.3 información sobre el gasto de consumo en relación con el ingreso de una muestra transversal de 30 familias. Suponga que postulamos que el gasto de consumo está relacionado linealmente con el ingreso pero que hay heteroscedasticidad en los datos. Postulamos además que la naturaleza de la heteroscedasticidad es como la de (11.5.10). En la tabla 11.3 presentamos también el reordenamiento necesario de los datos para aplicar la prueba.

Al eliminar las 4 observaciones del medio, las regresiones de MCO basadas en las primeras 13 observaciones y en las últimas 13, y sus sumas de cuadrados residuales asociadas se presentan a continuación (los errores estándar se indican entre paréntesis).

(*continúa*)

[19] Técnicamente, la **potencia** de la prueba depende de la forma de seleccionar *c*. En estadística, la *potencia* de una *prueba* se mide por la probabilidad de rechazar la hipótesis nula cuando ésta es falsa [es decir, 1 − Prob(error tipo II)]. Aquí la hipótesis nula es que las varianzas de los dos grupos son las mismas, es decir, que hay homoscedasticidad. Para mayor análisis, véase M. M. Ali y C. Giaccotto, "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model", *Journal of Econometrics,* vol. 26, 1984, pp. 355-373.

[20] George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lütkepohl y Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics,* John Wiley & Sons, Nueva York, 1982, p. 422.

**EJEMPLO 11.4** (*continuación*)

**TABLA 11.3** **Datos hipotéticos sobre el gasto de consumo $Y$($) y el ingreso $X$($) para ilustrar la prueba de Goldfeld-Quandt**

| | | Datos ordenados de acuerdo con los valores $X$ | | |
|---|---|---|---|---|
| $Y$ | $X$ | $Y$ | $X$ | |
| 55 | 80 | 55 | 80 | |
| 65 | 100 | 70 | 85 | |
| 70 | 85 | 75 | 90 | |
| 80 | 110 | 65 | 100 | |
| 79 | 120 | 74 | 105 | |
| 84 | 115 | 80 | 110 | |
| 98 | 130 | 84 | 115 | |
| 95 | 140 | 79 | 120 | |
| 90 | 125 | 90 | 125 | |
| 75 | 90 | 98 | 130 | |
| 74 | 105 | 95 | 140 | |
| 110 | 160 | 108 | 145 | |
| 113 | 150 | 113 | 150 | |
| 125 | 165 | 110 | 160 | 4 observaciones del medio |
| 108 | 145 | 125 | 165 | |
| 115 | 180 | 115 | 180 | |
| 140 | 225 | 130 | 185 | |
| 120 | 200 | 135 | 190 | |
| 145 | 240 | 120 | 200 | |
| 130 | 185 | 140 | 205 | |
| 152 | 220 | 144 | 210 | |
| 144 | 210 | 152 | 220 | |
| 175 | 245 | 140 | 225 | |
| 180 | 260 | 137 | 230 | |
| 135 | 190 | 145 | 240 | |
| 140 | 205 | 175 | 245 | |
| 178 | 265 | 189 | 250 | |
| 191 | 270 | 180 | 260 | |
| 137 | 230 | 178 | 265 | |
| 189 | 250 | 191 | 270 | |

Regresión basada en las primeras 13 observaciones:

$$\hat{Y}_i = 3.4094 + 0.6968X_i$$
$$(8.7049) \quad (0.0744) \qquad r^2 = 0.8887 \qquad \text{SCR}_1 = 377.17 \qquad \text{gl} = 11$$

Regresión basada en las últimas 13 observaciones:

$$\hat{Y}_i = -28.0272 + 0.7941X_i$$
$$(30.6421) \quad (0.1319) \qquad r^2 = 0.7681 \qquad \text{SCR}_2 = 1\,536.8 \qquad \text{gl} = 11$$

De estos resultados obtenemos

$$\lambda = \frac{\text{SCR}_2/\text{gl}}{\text{SCR}_1/\text{gl}} = \frac{1\,536.8/11}{377.17/11}$$
$$\lambda = 4.07$$

El valor $F$ crítico para 11 gl en el numerador y 11 gl en el denominador en el nivel de 5% es 2.82. Como el valor $F(= \lambda)$ estimado excede al valor crítico, podemos concluir que hay heteroscedasticidad en la varianza del error. Sin embargo, si el nivel de significancia se fija en 1%, no podemos rechazar el supuesto de homoscedasticidad. (¿Por qué?) Observe que el valor $p$ de la $\lambda$ observada es 0.014.

### *Prueba Breusch-Pagan-Godfrey*[21]

El éxito de la prueba de Goldfeld-Quandt depende no sólo del valor de $c$ (el número de observaciones centrales que se van a omitir), sino también de la identificación de la variable $X$ correcta que servirá de referencia para ordenar las observaciones. Esta limitación de la prueba se evita si consideramos la prueba Breusch-Pagan-Godfrey (BPG).

Para ilustrar esta prueba, considere el modelo de regresión lineal con $k$ variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{11.5.12}$$

Suponga que la varianza del error $\sigma_i^2$ se describe como

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi}) \tag{11.5.13}$$

es decir, $\sigma_i^2$ es algún tipo de función de las variables $Z$ no estocásticas; alguna de las $X$ o todas ellas pueden servir como $Z$. Específicamente, suponga que

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} \tag{11.5.14}$$

es decir, $\sigma_i^2$ es una función lineal de las $Z$. Si $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$, $\sigma_i^2 = \alpha_1$, que es una constante. Por consiguiente, para probar si $\sigma_i^2$ es homoscedástica, se puede probar la hipótesis de que $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$. Ésta es la idea básica de la prueba Breusch-Pagan. El procedimiento es el siguiente.

**Paso 1.** Estime (11.5.12) mediante MCO y obtenga los residuos $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$.

**Paso 2.** Obtenga $\tilde{\sigma}^2 = \sum \hat{u}_i^2/n$. Recuerde, del capítulo 4, que éste es el estimador de máxima verosimilitud (MV) de $\sigma^2$. [*Nota:* El estimador de MCO es $\sum \hat{u}_i^2/[n-k]$.]

**Paso 3.** Construya las variables $p_i$ definidas como

$$p_i = \hat{u}_i^2 \big/ \tilde{\sigma}^2$$

que es simplemente cada residuo elevado al cuadrado dividido entre $\tilde{\sigma}^2$.

**Paso 4.** Haga la regresión de los $p_i$ así construidos sobre las $Z$ como

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i \tag{11.5.15}$$

donde $v_i$ es el término de residuo para esta regresión.

**Paso 5.** Obtenga la SCE (suma de cuadrados explicada) de (11.5.15) y defina

$$\Theta = \frac{1}{2}(\text{SCE}) \tag{11.5.16}$$

Si suponemos que los $u_i$ están normalmente distribuidos, se demuestra que sí hay homoscedasticidad, y si el tamaño $n$ de la muestra aumenta indefinidamente, entonces

$$\Theta \underset{\text{asin}}{\sim} \chi^2_{m-1} \tag{11.5.17}$$

es decir, $\Theta$ sigue una distribución ji cuadrada con $(m-1)$ grados de libertad. (*Nota:* asin significa asintóticamente.)

---

[21] T. Breusch y A. Fagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica,* vol. 47, 1979, pp. 1287-1294. Véase también L. Godfrey, "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics,* vol. 8, 1978, pp. 227-236. Debido a la similitud, estas pruebas se conocen como pruebas de heteroscedasticidad Breusch-Pagan-Godfrey.

Por consiguiente, si en una aplicación el $\Theta$ $(= \chi^2)$ calculado excede al valor crítico $\chi^2$ en el nivel de significancia seleccionado, se rechaza la hipótesis de homoscedasticidad; de lo contrario, no se rechaza.

El lector puede preguntar por qué la prueba BPG elige $\frac{1}{2}$ SCE como estadístico de prueba. El razonamiento es un poco complicado y se deja para consultar en las referencias.[22]

**EJEMPLO 11.5**
*La prueba Breusch-Pagan-Godfrey (BPG)*

A manera de ejemplo, reconsidere la información (tabla 11.3) para ilustrar la prueba de heteroscedasticidad de Goldfeld-Quandt. Al efectuar la regresión de $Y$ sobre $X$, obtenemos lo siguiente:

**Paso 1.**

$$\begin{aligned} \hat{Y}_i &= 9.2903 + 0.6378X_i \\ \text{ee} &= (5.2314) \quad (0.0286) \qquad \text{SCR} = 2\,361.153 \qquad R^2 = 0.9466 \end{aligned} \tag{11.5.18}$$

**Paso 2.**

$$\tilde{\sigma}^2 = \sum \hat{u}_i^2/30 = 2\,361.153/30 = 78.7051$$

**Paso 3.** Divida los residuos $\hat{u}_i$ obtenidos de la regresión (11.5.18) entre 78.7051 para construir la variable $p_i$.

**Paso 4.** Si supone que las $p_i$ están relacionadas linealmente con $X_i$ $(= Z_i)$, como lo establece (11.5.14), obtenemos la regresión

$$\begin{aligned} \hat{p}_i &= -0.7426 + 0.0101X_i \\ \text{ee} &= (0.7529) \quad (0.0041) \qquad \text{SCE} = 10.4280 \qquad R^2 = 0.18 \end{aligned} \tag{11.5.19}$$

**Paso 5.**

$$\Theta = \frac{1}{2}(\text{SCE}) = 5.2140 \tag{11.5.20}$$

Con los supuestos de la prueba BPG, $\Theta$ en (11.5.20) sigue asintóticamente la distribución ji cuadrada con 1 gl. [*Nota:* sólo hay una regresora en (11.5.19).] Ahora, de la tabla ji cuadrada vemos que, para 1 gl, el valor crítico de ji cuadrada en 5% es 3.8414, y el valor $\chi^2$ crítico en 1% es 6.6349. Por tanto, el valor observado ji cuadrada de 5.2140 es significativo en el nivel de significancia de 5% pero no en el nivel de 1%. Por consiguiente, llegamos a la misma conclusión obtenida mediante la prueba Goldfeld-Quandt. Pero tenga en mente que, en estricto sentido, la prueba BPG es asintótica o de muestras grandes, y en el ejemplo presente, la muestra de 30 observaciones puede no ser una muestra grande. Debe señalarse también que, en muestras pequeñas, la prueba es sensible al supuesto de que las perturbaciones $u_i$ están normalmente distribuidas. Desde luego, podemos probar el supuesto de normalidad con las pruebas analizadas en el capítulo 5.[23]

### *Prueba general de heteroscedasticidad de White*

A diferencia de la prueba de Goldfeld-Quandt, que requiere reordenar las observaciones respecto de la variable $X$ que supuestamente ocasiona la heteroscedasticidad, o de la prueba BGP, sensible al supuesto de normalidad, la prueba general de heteroscedasticidad propuesta por White no se

[22] Véase Adrian C. Darnell, *A Dictionary of Econometrics,* Edward Elgar, Cheltenham, Reino Unido, 1994, pp. 178-179.

[23] Al respecto, véase R. Koenker, "A Note on Studentizing a Test for Heteroscedasticity", *Journal of Econometrics,* vol. 17, 1981, pp. 1180-1200.

apoya en el supuesto de normalidad y es fácil aplicarla.[24] Como ilustración de la idea básica, considere el siguiente modelo de regresión con tres variables (la generalización al modelo con $k$ variables es sencilla):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{11.5.21}$$

Para realizar la prueba de White se procede de la siguiente forma:

**Paso 1.** Dada la información, estime (11.5.21) y obtenga los residuos $\hat{u}_i$.

**Paso 2.** Efectúe la siguiente regresión (*auxiliar*):

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \qquad \textbf{(11.5.22)}^{25}$$

Es decir, con el cuadrado de los residuos de la regresión original se hace la regresión sobre las variables o regresoras $X$ originales, sobre sus valores al cuadrado y sobre el (los) producto(s) cruzado(s) de las regresoras. También pueden introducirse potencias más altas de las regresoras. Observe que hay un término constante en esta ecuación, aunque la regresión original puede o no contenerlo. Obtenga $R^2$ de esta regresión (auxiliar).

**Paso 3.** Según la hipótesis nula de que no hay heteroscedasticidad, puede demostrarse que el tamaño de la muestra ($n$) multiplicado por $R^2$ obtenido de la regresión auxiliar *asintóticamente* sigue la distribución ji cuadrada con gl igual al número de regresoras (sin el término constante) en la regresión auxiliar. Es decir,

$$n \cdot R^2 \underset{\text{asin}}{\sim} \chi^2_{\text{gl}} \tag{11.5.23}$$

donde los gl son iguales a los definidos antes. En el ejemplo, hay 5 gl porque hay 5 regresoras en la regresión auxiliar.

**Paso 4.** Si el valor ji cuadrada obtenido en (11.5.23) excede al valor ji cuadrada crítico en el nivel de significancia seleccionado, la conclusión es que hay heteroscedasticidad. Si éste no excede el valor ji cuadrada crítico, no hay heteroscedasticidad, lo cual quiere decir que en la regresión auxiliar (11.5.22), $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$ (véase la nota 25).

**EJEMPLO 11.6**
*Prueba de heteroscedasticidad de White*

Con información de corte transversal de 41 países, Stephen Lewis estimó el siguiente modelo de regresión:[26]

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \tag{11.5.24}$$

donde $Y$ = razón entre impuestos arancelarios (impuestos sobre importaciones y exportaciones) y ganancias totales del gobierno, $X_2$ = razón entre la suma de exportaciones e importaciones y el PNB, $X_3$ = PNB *per cápita*, y ln representa el logaritmo natural. Sus hipótesis fueron que $Y$ y $X_2$ estarían relacionadas positivamente (a mayor volumen de comercio exterior, mayor re-

(*continúa*)

[24] H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity", *Econometrica,* vol. 48, 1980, pp. 817-818.

[25] El supuesto de que la varianza del error $u_i$, $\sigma_i^2$, se relaciona funcionalmente con las regresoras, con sus valores al cuadrado y con sus productos cruzados está implícito en este procedimiento. Si todos los coeficientes parciales de pendiente en esta regresión son simultáneamente iguales a cero, entonces la varianza del error es homoscedástica e igual a la constante $\alpha_1$.

[26] Stephen R. Lewis, "Government Revenue from Foreign Trade", *Manchester School of Economics and Social Studies,* vol. 31, 1963, pp. 39-47.

**EJEMPLO 11.6** (*continuación*)

caudo arancelario), y que $Y$ y $X_3$ estarían negativamente relacionados (a medida que aumenta el ingreso, al gobierno se le facilita recaudar impuestos directos —es decir, el impuesto sobre la renta— que depende de los impuestos sobre el comercio exterior).

Los resultados empíricos apoyaron las hipótesis. Para el propósito, el punto importante es averiguar si hay heteroscedasticidad en los datos. Como los datos son de corte transversal e implican una heterogeneidad de países, podemos esperar *a priori* heteroscedasticidad en la varianza del error. Con la prueba de heteroscedasticidad de White en los residuos obtenidos de la regresión (11.5.24) se obtuvieron los siguientes resultados:[27]

$$\begin{aligned}\widehat{\hat{u}_i^2} = &-5.8417 + 2.5629 \ln \text{Comercio}_i + 0.6918 \ln \text{PNB}_i \\ &-0.4081(\ln \text{Comercio}_i)^2 - 0.0491(\ln \text{PNB}_i)^2 \\ &+0.0015(\ln \text{Comercio}_i)(\ln \text{PNB}_i) \qquad R^2 = 0.1148\end{aligned} \tag{11.5.25}$$

*Nota:* Los errores estándar no están dados porque no son pertinentes para nuestro propósito.

Ahora, $n \cdot R^2 = 41(0.1148) = 4.7068$, que tiene, asintóticamente, una distribución ji cuadrada con 5 gl (¿por qué?). El valor ji cuadrada crítico en 5% para 5 gl es 11.0705, el valor crítico en 10% es 9.2363, y el valor crítico en 25% es 6.62568. Para todos los fines prácticos, podemos concluir, con base en la prueba de White, que no hay heteroscedasticidad.

Conviene hacer un comentario relacionado con la prueba de White. Si un modelo tiene muchas regresoras, la introducción de todas las regresoras, de sus términos elevados al cuadrado (o a potencias más elevadas) y de sus productos cruzados pueden consumir grados de libertad rápidamente. Por consiguiente, se debe tener cautela con esta prueba.[28]

En los casos en que el estadístico de prueba de White es significativo estadísticamente, la heteroscedasticidad puede no necesariamente ser la causa, sino los errores de especificación, los cuales veremos en mayor detalle en el capítulo 13 (recuerde el punto 5 de la sección 11.1). En otras palabras, **la prueba de White puede ser una prueba de heteroscedasticidad (pura), de error de especificación o de ambos**. Se argumenta que, si no están presentes términos con productos cruzados en el procedimiento de prueba de White, esto constituye una prueba de heteroscedasticidad pura. Si existen tales términos, es una prueba de heteroscedasticidad y de sesgo de especificación.[29]

### *Otras pruebas de heteroscedasticidad*

Hay muchas otras pruebas de heteroscedasticidad, cada una con supuestos determinados. El lector interesado puede consultar las referencias.[30] Mencionamos sólo una de estas pruebas debido a su simplicidad. Es la **prueba de Koenker-Basset (KB)**. Al igual que las pruebas Park, Breusch-Pagan-Godfrey y la de White, la prueba KB se basa en los residuos al cuadrado, $\hat{u}_i^2$, pero en vez de hacer la regresión sobre una o más regresoras, se efectúa la regresión de los residuos al cuadrado sobre los valores estimados de la regresora al cuadrado. De manera específica, si el modelo original es:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{11.5.26}$$

[27] Estos resultados, con cambio en la notación, se reproducen de William F. Lott y Subhash C. Ray, *Applied Econometrics: Problems with Data Sets, Instructor's Manual*, capítulo 22, pp. 137-140.

[28] A veces la prueba se puede modificar para conservar los grados de libertad. Consulte el ejercicio 11.18.

[29] Véase Richard Harris, *Using Cointegration Analysis in Econometrics Modelling,* Prentice Hall & Harvester Wheatsheaf, Reino Unido, 1995, p. 68.

[30] Véase M.J. Harrison y B.E. McCabe, "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals", *Journal of the American Statistical Association,* vol. 74, 1979, pp. 494-499; J. Szroeter, "A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models", *Econometrica,* vol. 46, 1978, pp. 1311-1327; M.A. Evans y M.L. King, "A Further Class of Tests for Heteroscedasticity", *Journal of Econometrics,* vol. 37, 1988, pp. 265-276; y R. Koenker y G. Bassett, "Robust Tests for Heteroscedasticity Based on Regression Quantiles", *Econometrica,* vol. 50, 1982, pp. 43-61.

se estima este modelo, se obtiene $\hat{u}_i$ de dicho modelo y luego se calcula

$$\hat{u}_i^2 = \alpha_1 + \alpha_2(\hat{Y}_i)^2 + v_i \quad \textbf{(11.5.27)}$$

donde $\hat{Y}_i$ son los valores estimados del modelo (11.5.26). La hipótesis nula es que $\alpha_2 = 0$. Si no se rechaza, se puede concluir que no existe heteroscedasticidad. La hipótesis nula se prueba con las pruebas $t$ o $F$ usuales. (Observe que $F_{1,k} = t_k{}^2$.) Si el modelo (11.5.26) es doble logaritmo, se lleva a cabo la regresión de los residuos al cuadrado sobre (log $\hat{Y}_i)^2$. Otra ventaja de la prueba KB es que es aplicable aunque el término de error en el modelo original (11.5.26) no esté normalmente distribuido. Si aplica la prueba KB al ejemplo 11.1, descubrirá que el coeficiente de la pendiente en la regresión de los residuos cuadrados obtenida de (11.5.3) sobre el $\hat{Y}_i^2$ estimado a partir de (11.5.3) no es estadísticamente distinto de cero, por lo que se refuerza la prueba de Park. Este resultado no debe sorprender, pues en estos momentos sólo se tiene una sola regresora. No obstante, la prueba KB es aplicable si hay una o muchas regresoras.

*Nota sobre las pruebas de heteroscedasticidad*

Ya analizamos varias pruebas de heteroscedasticidad en esta sección. Pero, ¿cómo decidir cuál es la mejor? No es una pregunta fácil, pues estas pruebas se basan en supuestos diversos. Al comparar las pruebas, es necesario prestar atención al tamaño (o nivel de significancia), potencia (la probabilidad de rechazar una hipótesis falsa) y sensibilidad a los valores atípicos.

Ya señalamos algunas limitaciones de la prueba de heteroscedasticidad de White, que es popular y fácil de aplicar. Como resultado de estas limitaciones, tal vez tenga poca potencia en relación con las opciones. Además, la prueba no sirve para identificar los factores o variables que causan heteroscedasticidad.

Asimismo, la prueba de Breusch-Pagan-Godfrey es sensible al supuesto de normalidad. En contraste, la prueba de Koenker-Bassett no se basa en el supuesto de normalidad y, en consecuencia, puede ser más potente.[31] En la prueba de Goldfeld-Quandt, si se omiten muchas observaciones, puede disminuir la potencia de la prueba.

Está fuera del ámbito de este texto proporcionar un análisis comparativo de las diferentes pruebas de heteroscedasticidad. Sin embargo, el lector interesado puede consultar el artículo de John Lyon y Chin-Ling Tsai para darse una idea de los puntos fuertes y débiles de las diversas pruebas de heteroscedasticidad.[32]

## 11.6 Medidas correctivas

Como vimos, la heteroscedasticidad no destruye las propiedades de insesgamiento y consistencia de los estimadores de MCO; sin embargo, éstos ya no son eficientes, ni siquiera asintóticamente (es decir, en muestras grandes). Esta falta de eficiencia resta credibilidad a los procedimientos habituales de pruebas de hipótesis. Por consiguiente, es necesario introducir medidas correctivas. Existen dos enfoques para remediar el problema de heteroscedasticidad: cuando se conoce $\sigma_i^2$ y cuando no se conoce $\sigma_i^2$.

### Cuando se conoce $\sigma_i^2$: método de los mínimos cuadrados ponderados

Como vimos en la sección 11.3, si se conoce $\sigma_i^2$, el método más directo de corregir la heteroscedasticidad es con los mínimos cuadrados ponderados, pues los estimadores obtenidos mediante este método son MELI.

---

[31] Para detalles, véase William H. Green, *Econometric Analysis,* 6a. ed., Pearson/Prentice-Hall, Nueva Jersey, 2008, pp. 165-167.

[32] Véase su artículo "A Comparison of Tests of Heteroscedasticity", *The Statistician,* vol. 45, núm. 3, 1996, pp. 337-349.

**EJEMPLO 11.7**
*Ilustración del método de los mínimos cuadrados ponderados*

Para ilustrar el método, suponga que deseamos estudiar la relación entre la remuneración salarial y el tamaño de la planta laboral de los datos de la tabla 11.1. Por simplicidad, se mide el tamaño de la planta laboral con las siguientes categorías: 1 (1-4 empleados), 2 (5-9 empleados), . . . , 9 (1 000-2 499 empleados), aunque también se puede medir con el punto medio de las diversas clases de empleados en la tabla.

Ahora, sea $Y$ la remuneración salarial promedio por empleado ($) y $X$ el tamaño de empleados. Efectuamos la siguiente regresión [véase la ecuación (11.3.6)]:

$$Y_i/\sigma_i = \hat{\beta}_1^*(1/\sigma_i) + \hat{\beta}_2^*(X_i/\sigma_i) + (\hat{u}_i/\sigma_i) \tag{11.6.1}$$

donde $\sigma_i$ son las desviaciones estándar de los salarios como aparecen en la tabla 11.1. Los datos simples necesarios para efectuar esta regresión están en la tabla 11.4.

**TABLA 11.4**
**Ilustración de una regresión de mínimos cuadrados ponderados**

Fuente: La información sobre $Y$ y $\sigma_i$ (desviación estándar de la remuneración salarial) corresponde a la tabla 11.1. El tamaño de la planta laboral: 1 = 1-4 empleados, 2 = 5-9 empleados, etc. Los últimos datos son también de la tabla 11.1.

| Remuneración, $Y$ | Tamaño de la planta laboral, $X$ | $\sigma_i$ | $Y_i/\sigma_i$ | $X_i/\sigma_i$ |
|---|---|---|---|---|
| 3 396 | 1 | 742.2 | 4.5664 | 0.0013 |
| 3 787 | 2 | 851.4 | 4.4480 | 0.0023 |
| 4 013 | 3 | 727.8 | 5.5139 | 0.0041 |
| 4 104 | 4 | 805.06 | 5.0978 | 0.0050 |
| 4 146 | 5 | 929.9 | 4.4585 | 0.0054 |
| 4 241 | 6 | 1 080.6 | 3.9247 | 0.0055 |
| 4 387 | 7 | 1 241.2 | 3.5288 | 0.0056 |
| 4 538 | 8 | 1 307.7 | 3.4702 | 0.0061 |
| 4 843 | 9 | 1 110.7 | 4.3532 | 0.0081 |

*Nota:* En la regresión (11.6.2), la variable dependiente es $(Y_i/\sigma_i)$, y las independientes, $(1/\sigma_i)$ y $(X_i/\sigma_i)$.

Antes de proseguir con el análisis de los resultados de la regresión, observe que (11.6.1) no tiene término de intercepto (¿por qué?). Por consiguiente, debemos utilizar el modelo de regresión a través del origen para estimar $\beta_1^*$ y $\beta_2^*$, tema analizado en el capítulo 6. Pero, hoy en día, la mayoría de los paquetes de computación dan la opción de suprimir el término de intercepto (*Minitab* o *EViews*, por ejemplo). Observe también otra característica interesante de (11.6.1): tiene dos variables explicativas, $(1/\sigma_i)$ y $(X_i/\sigma_i)$, mientras que si utilizáramos MCO, la regresión del salario sobre el tamaño de la planta laboral tendría una sola variable explicativa, $X_i$. (¿Por qué?)

Los resultados de la regresión de MCP son los siguientes:

$$\begin{aligned}\widehat{(Y_i/\sigma_i)} &= 3\,406.639(1/\sigma_i) + 154.153(X_i/\sigma_i)\\ &\quad\;(80.983) \qquad\qquad (16.959)\\ t &= \;(42.066) \qquad\qquad\; (9.090)\\ &\quad R^2 = 0.9993^{33}\end{aligned} \tag{11.6.2}$$

Para su comparación, presentamos a continuación los resultados de la regresión de MCO usual o no ponderada:

$$\begin{aligned}\hat{Y}_i &= 3\,417.833 + 148.767\,X_i\\ &\quad\;(81.136) \quad (14.418)\\ t &= \;(42.125) \quad (10.318) \qquad R^2 = 0.9383\end{aligned} \tag{11.6.3}$$

En el ejercicio 11.7 se le pide comparar estas dos regresiones.

[33] Como se señaló en la nota 3 del capítulo 6, $R^2$ de la regresión a través del origen no es directamente comparable con $R^2$ del modelo con presencia del intercepto. El valor $R^2$ de 0.9993 obtenido tiene en cuenta esta diferencia. (Véanse los diversos paquetes de cómputo para mayores detalles sobre la forma de corregir $R^2$ para tener en cuenta la ausencia del término del intercepto. Véase también el apéndice 6A, sección 6A1.)

## Cuando no se conoce $\sigma_i^2$

Como ya vimos, si se conocen las verdaderas $\sigma_i^2$, podemos utilizar el método de MCP para obtener estimadores MELI. Como pocas veces se conocen las verdaderas $\sigma_i^2$, ¿existe alguna forma de obtener estimaciones *consistentes* (en el sentido estadístico) de las varianzas y covarianzas de los estimadores de MCO aunque haya heteroscedasticidad? La respuesta es sí.

### *Varianzas y errores estándar consistentes con heteroscedasticidad de White*

White demostró que esta estimación puede realizarse de forma que las inferencias estadísticas sean *asintóticamente* válidas (es decir, para muestras grandes) sobre los verdaderos valores de los parámetros.[34] No presentaremos aquí los detalles matemáticos porque no están al alcance de este libro. En el apéndice 11A.4 se detalla el procedimiento de White. Sin embargo, en la actualidad hay diversos paquetes de computación que presentan varianzas y errores estándar con la corrección de heteroscedasticidad de White en forma simultánea con las varianzas y los errores estándar de MCO usuales.[35] A propósito, los errores estándar de White corregidos mediante heteroscedasticidad también se conocen como **errores estándar robustos**.

**EJEMPLO 11.8**
*Ilustración del procedimiento de White*

Como ejemplo, citamos los siguientes resultados obtenidos por Greene:[36]

$$
\begin{aligned}
\hat{Y}_i &= \ \ 832.91 \ - \ 1\,834.2\,(\text{Ingreso}) + 1\,587.04\,(\text{Ingreso})^2 \\
\text{ee de MCO} &= (327.3) \qquad (829.0) \qquad\qquad (519.1) \\
t &= \ \ (2.54) \qquad\ (2.21) \qquad\qquad\ (3.06) \qquad\qquad (11.6.4) \\
\text{ee de White} &= (460.9) \quad\ (1\,243.0) \qquad\quad (830.0) \\
t &= \ \ (1.81) \qquad (-1.48) \qquad\qquad (1.91)
\end{aligned}
$$

donde $Y$ = gasto *per cápita* en escuelas públicas por estado en 1979 e Ingreso = ingreso *per cápita* por estado en 1979. La muestra consistió en 50 estados más Washington, D.C.

Como demuestran los resultados anteriores, los errores estándar corregidos por heteroscedasticidad (de White) resultan considerablemente más grandes que los errores estándar de MCO, y, por consiguiente, los valores $t$ estimados son mucho menores que los obtenidos por MCO. Con base en estos últimos, ambas regresoras son estadísticamente significativas en el nivel de 5%, mientras que con base en los estimadores de White, no lo son. Sin embargo, cabe señalar que los errores estándar corregidos por heteroscedasticidad de White pueden ser más grandes o más pequeños que los errores estándar sin corregir.

Como los estimadores de las varianzas consistentes con heteroscedasticidad de White están disponibles ahora en paquetes de computación para regresión, se recomienda que el lector los reporte. Como recomiendan Wallace y Silver:

> En términos generales, quizá sea buena idea utilizar la opción WHITE [disponible en los programas de regresión] sistemáticamente, tal vez comparar estos resultados con los resultados de MCO regulares como forma de verificar si la heteroscedasticidad es un problema grave en un conjunto particular de datos.[37]

### *Supuestos razonables sobre el patrón de heteroscedasticidad*

Una desventaja del procedimiento de White, además de ser de muestras grandes, es que los estimadores obtenidos por este medio pueden no ser tan eficientes como los obtenidos por métodos

[34] Véase H. White, *op. cit.*

[35] Más técnicamente, se conocen como **estimadores de la matriz de covarianzas consistentes con heteroscedasticidad**: para abreviar, **EMCCH**.

[36] William H. Greene, *Econometric Analysis,* 2a. ed., Macmillan, Nueva York, 1993, p. 385.

[37] T. Dudley Wallace y J. Lew Silver, *Econometrics: An Introduction,* Addison-Wesley, Reading, Massachusetts, 1988, p. 265.

que transforman la información para reflejar tipos específicos de heteroscedasticidad. Para ilustrar esto, recordemos el modelo de regresión con dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Consideraremos ahora diversos supuestos sobre el patrón de heteroscedasticidad.

**SUPUESTO 1** La varianza del error es proporcional a $X_i^2$:

$$E\left(u_i^2\right) = \sigma^2 X_i^2 \qquad \textbf{(11.6.5)}^{38}$$

Si, por razones de "especulación", por los métodos gráficos, o por los métodos Park y Glejser, se cree que la varianza de $u_i$ es proporcional al cuadrado de la variable explicativa $X$ (figura 11.10), se puede transformar el modelo original de la siguiente manera. Divida el modelo original entre $X_i$:

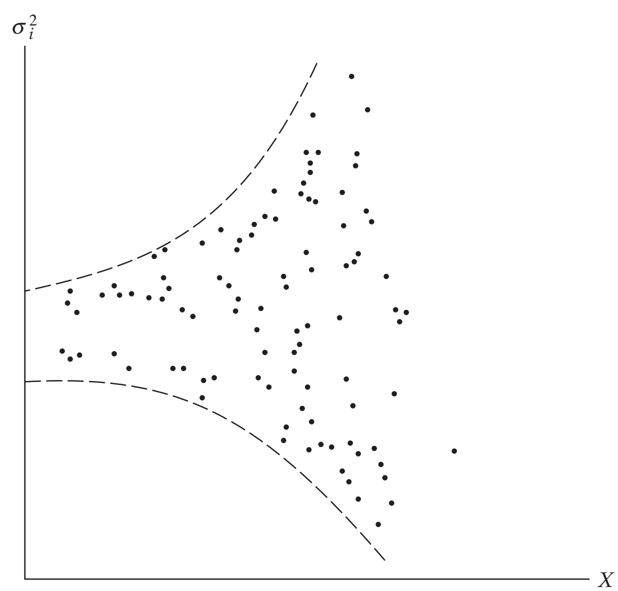
$$\begin{aligned} \frac{Y_i}{X_i} &= \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} \\ &= \beta_1 \frac{1}{X_i} + \beta_2 + v_i \end{aligned} \qquad \textbf{(11.6.6)}$$

donde $v_i$ es el término de perturbación transformado, igual a $u_i/X_i$. Ahora, es fácil verificar que

$$\begin{aligned} E\left(v_i^2\right) &= E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E\left(u_i^2\right) \\ &= \sigma^2 \qquad \text{utilizando (11.6.5)} \end{aligned}$$

Por tanto, la varianza de $v_i$ es ahora homoscedástica y podemos aplicar MCO a la ecuación transformada (11.6.6), con la regresión $Y_i/X_i$ sobre $1/X_i$.

**FIGURA 11.10**
Varianza del error proporcional a $X^2$.




[38] Recuerde que ya tratamos este supuesto en el análisis de la prueba Goldfeld-Quandt.

Observe que, en la regresión transformada, el término del intercepto $\beta_2$ es el coeficiente de pendiente en la ecuación original, y el coeficiente de la pendiente $\beta_1$ es el término del intercepto en el modelo original. Por consiguiente, para retornar al modelo original tenemos que multiplicar la estimación (11.6.6) por $X_i$. Una aplicación de esta transformación está en el ejercicio 11.20.

**SUPUESTO 2** La varianza del error es proporcional a $X_i$. La **transformación de raíz cuadrada**:

$$E\left(u_i^2\right) = \sigma^2 X_i \tag{11.6.7}$$

Si se cree que la varianza de $u_i$, en lugar de ser proporcional al cuadrado $X_i$, es proporcional a la misma $X_i$, entonces el modelo original se transforma de la siguiente manera (figura 11.11):

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}} &= \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ &= \beta_1\frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + v_i\end{aligned} \tag{11.6.8}$$

donde $v_i = u_i/\sqrt{X_i}$ y donde $X_i > 0$.

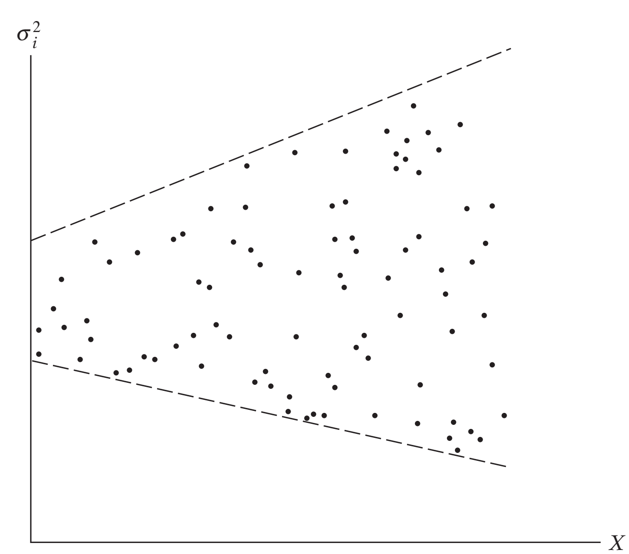
Con el supuesto 2 se verifica fácilmente que $E(v_i^2) = \sigma^2$, una situación homoscedástica. Por consiguiente, se puede aplicar MCO a (11.6.8), con la regresión de $Y_i/\sqrt{X_i}$ sobre $1/\sqrt{X_i}$ y $\sqrt{X_i}$.

Observe una característica importante del modelo transformado: no tiene término de intercepto. Por consiguiente, será necesario el modelo de regresión a través del origen para estimar $\beta_1$ y $\beta_2$. Tras efectuar la regresión (11.6.8), retornamos al modelo original con tan sólo multiplicar (11.6.8) por $\sqrt{X_i}$.

Un caso interesante es el modelo de intercepto cero, es decir, $Y_i = \beta_2 X_i + u_i$. En este caso, la ecuación (11.6.8) se convierte en:

$$\frac{Y_i}{\sqrt{X_i}} = \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \tag{11.6.8a}$$

**FIGURA 11.11**
Varianza del error proporcional a *X*.



Y se puede demostrar que

$$\hat{\beta}_2 = \frac{\bar{Y}}{\bar{X}} \tag{11.6.8b}$$

Es decir, el estimador de mínimos cuadrados ponderados es simplemente la razón de las medias de las variables dependiente y explicativa. (Para probar la ecuación [11.6.8b], se aplica la fórmula de regresión a través del origen dada en la ecuación [6.1.6].)

**SUPUESTO 3** La varianza del error es proporcional al cuadrado del valor medio de *Y*.

$$E\left(u_i^2\right) = \sigma^2[E(Y_i)]^2 \tag{11.6.9}$$

La ecuación (11.6.9) postula que la varianza de $u_i$ es proporcional al cuadrado del valor esperado de $Y$ (figura 11.8*e*). Ahora,

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

Por consiguiente, si transformamos la ecuación original de la siguiente manera,

$$\begin{aligned} \frac{Y_i}{E(Y_i)} &= \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} \\ &= \beta_1 \left(\frac{1}{E(Y_i)}\right) + \beta_2 \frac{X_i}{E(Y_i)} + v_i \end{aligned} \tag{11.6.10}$$

donde $v_i = u_i/E(Y_i)$, se ve que $E(v_i^2) = \sigma^2$; es decir, las perturbaciones $v_i$ son homoscedásticas. Por tanto, es la regresión (11.6.10) la que satisfará el supuesto de homoscedasticidad del modelo clásico de regresión lineal.

La transformación (11.6.10), sin embargo, no funciona, porque $E(Y_i)$ depende de $\beta_1$ y $\beta_2$, los cuales no se conocen. Por supuesto, se conoce $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$, que es un estimador de $E(Y_i)$. Por consiguiente, podemos proceder en dos etapas: primero, efectuamos la regresión de MCO usual sin considerar el problema de heteroscedasticidad y obtenemos $\hat{Y}_i$. Luego, con el $\hat{Y}_i$ estimado, transformamos el modelo de la siguiente manera:

$$\frac{Y_i}{\hat{Y}_i} = \beta_1 \left(\frac{1}{\hat{Y}_i}\right) + \beta_2 \left(\frac{X_i}{\hat{Y}_i}\right) + v_i \tag{11.6.11}$$

donde $v_i = (u_i/\hat{Y}_i)$. En el paso 2 efectuamos la regresión (11.6.11). Aunque $\hat{Y}_i$ no es exactamente $E(Y_i)$, *estos estimadores son consistentes*; es decir, a medida que el tamaño de la muestra aumenta indefinidamente, éstos convergen hacia el verdadero $E(Y_i)$. Por tanto, la transformación (11.6.11) tendrá un desempeño adecuado en la práctica si el tamaño de la muestra es razonablemente grande.

**SUPUESTO 4** Una transformación logarítmica como

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{11.6.12}$$

con gran frecuencia reduce la heteroscedasticidad cuando se compara con la regresión $Y_i = \beta_1 + \beta_2 X_i + u_i$.

Este resultado surge porque la transformación logarítmica comprime las escalas en las cuales se miden las variables, y por tanto reduce una diferencia entre dos valores de diez veces a una diferencia de dos veces. Así, el número 80 es diez veces el número 8, pero el ln 80 (= 4.3280) es cerca de dos veces tan grande como ln 8 (= 2.0794).

Otra ventaja de la transformación logarítmica es que el coeficiente de pendiente $\beta_2$ mide la elasticidad de $Y$ respecto de $X$, es decir, el cambio porcentual en $Y$ ante un cambio porcentual en $X$. Por ejemplo, si $Y$ es el consumo y $X$ el ingreso, $\beta_2$ en (11.6.12) mide la elasticidad del ingreso, mientras que en el modelo original, $\beta_2$ sólo mide la tasa de cambio del consumo medio por cambio unitario del ingreso. Ésta es una razón por la cual los modelos logarítmicos son muy populares en la econometría empírica. (En el ejercicio 11.4 se aprecian algunos problemas asociados a la transformación logarítmica.)

Para concluir la exposición sobre medidas correctivas, de nuevo destacamos que todas las transformaciones analizadas antes son *ad hoc*; en esencia, especulamos sobre la naturaleza de $\sigma_i^2$. El que las transformaciones estudiadas en verdad funcionen depende de la naturaleza del problema y de la gravedad de la heteroscedasticidad. Hay otros problemas con las transformaciones que deben tomarse en cuenta:

1. Cuando se va más allá del modelo con dos variables tal vez no se conozca *a priori* la variable $X$ que debe seleccionarse para transformar los datos.[39]

2. La transformación logarítmica como se analiza en el supuesto 4 no es aplicable si algún valor de $Y$ o de $X$ es cero o negativo.[40]

3. Además, hay un problema de **correlación espuria**. Este término, atribuido a Karl Pearson, se refiere a la situación de correlación entre las razones de variables, aunque las variables originales no estén correlacionadas o sean aleatorias.[41] Así, en el modelo $Y_i = \beta_1 + \beta_2 X_i + u_i$, $Y$ y $X$ pueden no estar correlacionados, pero en el modelo transformado $Y_i/X_i = \beta_1(1/X_i) + \beta_2$, a menudo se encuentra que $Y_i/X_i$ y $1/X_i$ sí lo están.

4. Cuando las $\sigma_i^2$ no se conocen directamente y se estiman a partir de una o más de las transformaciones ya analizadas, todos los procedimientos de prueba $t$, $F$, etc., son, *en estricto sentido, válidos sólo para muestras grandes*. Por consiguiente, se debe tener cuidado al interpretar resultados fundados en las diversas transformaciones cuando las muestras son pequeñas o finitas.[42]

## 11.7 Ejemplos para concluir

Para concluir el análisis de heteroscedasticidad presentamos tres ejemplos que ilustran los principales temas de este capítulo.

[39] Sin embargo, como un asunto práctico, se puede graficar $\hat{u}_i^2$ contra cada una de las variables y decidir cuál variable $X$ puede utilizarse para transformar la información (figura 11.9).

[40] Algunas veces podemos utilizar ln $(Y_i + k)$ o ln $(X_i + k)$, donde $k$ es un número positivo seleccionado de forma que todos los valores de $Y$ y $X$ se tornan positivos.

[41] Por ejemplo, si $X_1$, $X_2$ y $X_3$ están mutuamente no correlacionadas, $r_{12} = r_{13} = r_{23} = 0$ y se encuentra que las razones (sus valores) $X_1/X_3$ y $X_2/X_3$ están correlacionadas, entonces hay correlación espuria. "En términos más generales, la correlación se describe como espuria si es inducida por el método de manejo de datos y no está presente en la información original." M. G. Kendall y W. R. Buckland, *A Dictionary of Statistical Terms,* Hafner, Nueva York, 1972, p. 143.

[42] Para mayores detalles, véase George G. Judge *et al., op. cit.,* sección 14.4, pp. 415-420.

**EJEMPLO 11.9**
*De nuevo la mortalidad infantil*

Regresemos al ejemplo de mortalidad infantil que hemos analizado en diversas ocasiones. De los datos de 64 países obtuvimos los resultados de la regresión mostrados en la ecuación (8.1.4). En vista de que los datos son transversales e implican diferentes países con distintas experiencias sobre mortalidad infantil, parece muy probable que haya heteroscedasticidad. Para descubrirlo, primero tomaremos en cuenta los residuos obtenidos de la ecuación (8.1.4). Tales residuos se grafican en la figura 11.12. De acuerdo con dicha gráfica, al parecer los residuos no muestran ningún patrón específico que sugiera heteroscedasticidad. No obstante, las apariencias engañan, así que aplicaremos las pruebas de Park, Glejser y White para descubrir alguna evidencia de heteroscedasticidad.

**Prueba de Park.** Como hay dos regresoras, PIB y TAM, hacemos la regresión de los residuos al cuadrado a partir de la regresión (8.1.4) sobre cualquiera de esas variables. O podemos efectuar la regresión de ellos sobre los valores estimados de MI ($= \widehat{\text{MI}}$) a partir de la regresión (8.1.4). Con la regresión anterior obtenemos los siguientes resultados:

$$\begin{aligned} \widehat{\hat{u}_i^2} &= 854.4006 \;+\; 5.7016\,\widehat{\text{MI}}_i \\ t &= \quad (1.2010) \qquad (1.2428) \qquad\qquad r^2 = 0.024 \end{aligned} \tag{11.7.1}$$

*Nota:* $\hat{u}_i$ son los residuos obtenidos de la regresión (8.1.4) y $\widehat{\text{MI}}$ son los valores estimados de MI a partir de la regresión (8.1.4).

Como muestra esta regresión, no existe ninguna relación sistemática entre el cuadrado de los residuos y los valores estimados de MI (¿por qué?), lo cual sugiere que el supuesto sobre la homoscedasticidad puede ser válido. A propósito, si hacemos la regresión del logaritmo de los residuos al cuadrado sobre el logaritmo de $\widehat{\text{MI}}$, no cambia la conclusión.

**Prueba de Glejser.** Los valores absolutos de los residuos obtenidos a partir de (8.1.4), cuando se hizo la regresión de los valores estimados de MI a partir de esa misma regresión, dieron estos resultados:

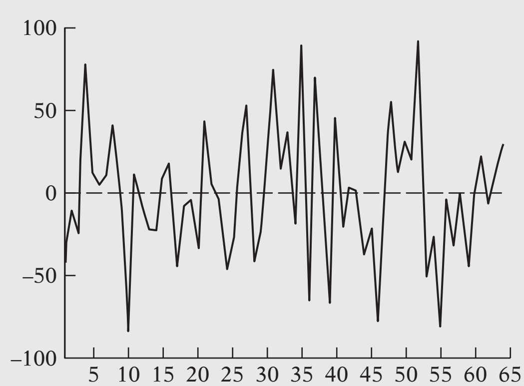
$$\begin{aligned} \widehat{|\hat{u}_i|} &= 22.3127 \;+\; 0.0646\,\widehat{\text{MI}}_i \\ t &= (2.8086) \qquad (1.2622) \qquad\qquad r^2 = 0.0250 \end{aligned} \tag{11.7.2}$$

De nuevo, no existe una gran relación sistemática entre los valores absolutos de los residuos y los valores estimados de MI, pues el valor *t* del coeficiente de la pendiente no es estadísticamente significativo.

**Prueba de White.** Al aplicar la prueba de White para la heteroscedasticidad con y sin los términos de los productos cruzados no hallamos ninguna evidencia de heteroscedasticidad. También volvimos a estimar (8.1.4) para obtener los errores estándar consistentes con heteroscedasticidad de White y los valores *t*, pero los resultados fueron muy parecidos a los de la ecuación (8.1.4), lo cual no sorprende en vista de lo arrojado por las diversas pruebas de heteroscedasticidad anteriores.

En resumen, parece que la regresión de mortalidad infantil (8.1.4) no experimenta heteroscedasticidad alguna.

**FIGURA 11.12**
Residuos de la regresión (8.1.4).



**EJEMPLO 11.10**
*Gastos en investigación y desarrollo (IyD), ventas y utilidades de 14 sectores industriales de Estados Unidos, 2005*

En la tabla 11.5 se presentan datos sobre el gasto en investigación y desarrollo (IyD), ventas y utilidades de 14 sectores industriales en Estados Unidos (todas las cifras se expresan en millones de dólares). Como los datos transversales presentados en la tabla 11.5 son muy heterogéneos, en una regresión de IyD sobre las ventas es probable la presencia de heteroscedasticidad. Los resultados obtenidos de la regresión fueron los siguientes:

$$\begin{aligned} \widehat{\text{IyD}}_i &= 1\,338 \quad + \quad 0.0437\ \text{Ventas}_i \\ \text{ee} &= (5\,015) \qquad (0.0277) \\ t &= (0.27) \qquad (1.58) \qquad r^2 = 0.172 \end{aligned} \tag{11.7.3}$$

No sorprende una relación positiva entre IyD y ventas, aunque no sea estadísticamente significativa en los niveles tradicionales.

**TABLA 11.5**
**Ventas y empleo para empresas que realizan IyD industrial en Estados Unidos, por industria, 2005 (los valores se expresan en millones de dólares)**

Fuente: National Science Foundation, Division of Science Resources Statistics, Survey of Industrial Research and Development, 2005, y U.S. Census Bureau Annual Survey of Manufacturers, 2005.

| Industria | Ventas | IyD | Utilidades |
|---|---|---|---|
| 1 Alimentos | 374 342 | 2 716 | 234 662 |
| 2 Textiles, ropa y cuero | 51 639 | 816 | 53 510 |
| 3 Productos químicos básicos | 109 899 | 2 277 | 75 168 |
| 4 Resinas, caucho sintético, fibras y filamentos | 132 934 | 2 294 | 34 645 |
| 5 Productos farmacéuticos y medicamentos | 273 377 | 34 839 | 127 639 |
| 6 Productos de plástico y caucho | 90 176 | 1 760 | 96 162 |
| 7 Productos metálicos fabricados | 174 165 | 1 375 | 155 801 |
| 8 Maquinaria | 230 941 | 8 531 | 143 472 |
| 9 Computadoras y equipo periférico | 91 010 | 4 955 | 34 004 |
| 10 Semiconductores y otros componentes electrónicos | 176 054 | 18 724 | 81 317 |
| 11 Instrumentos de navegación, medición, electromédicos y de control | 118 648 | 15 204 | 73 258 |
| 12 Equipo eléctrico, aparatos electrodomésticos y componentes | 101 398 | 2 424 | 54 742 |
| 13 Productos y partes aeroespaciales | 227 271 | 15 005 | 72 090 |
| 14 Equipo y suministros médicos | 56 661 | 4 374 | 52 443 |

Para verificar si la regresión (11.7.3) experimenta heteroscedasticidad, obtuvimos los residuos $\hat{u}_i$ y los residuos al cuadrado $\hat{u}_i^2$ del modelo, y se graficaron respecto de las ventas, como se muestra en la figura 11.13. Según esta figura, parece que existe un patrón sistemático entre los residuos y los cuadrados de los residuos y las ventas, lo cual indica heteroscedasticidad. Para probar esto de manera formal, utilizamos las pruebas de Park, Glejser y White, las cuales dieron los siguientes resultados:

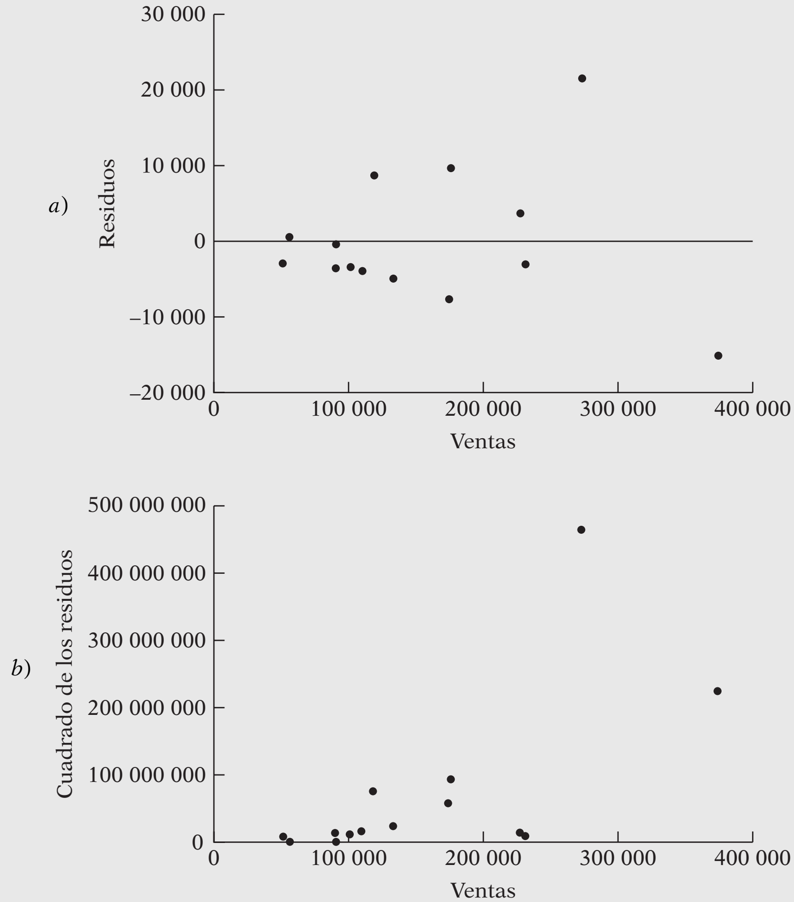
**Prueba de Park**

$$\begin{aligned} \widehat{\hat{u}_i^2} &= -72\,493\,719 \quad + \quad 916.1\ \text{Ventas}_i \\ \text{ee} &= (54\,940\,238) \qquad (303.9) \\ t &= (-1.32) \qquad (3.01) \qquad r^2 = 0.431 \end{aligned} \tag{11.7.4}$$

La prueba de Park indica una relación positiva estadísticamente significativa entre el cuadrado de los residuos y las ventas.

**EJEMPLO 11.10** (*continuación*)

FIGURA 11.13 Residuos *a*) y cuadrado de los residuos *b*) sobre ventas.




**Prueba de Glejser**

$$\widehat{|\hat{u}_i|} = -1\,003 + 0.04639 \text{ Ventas}_i$$
$$\text{ee} = (2\,316) \quad (0.0128) \tag{11.7.5}$$
$$t = (-0.43) \quad (3.62) \qquad r^2 = 0.522$$

La prueba de Glejser también indica una relación sistemática entre los valores absolutos de los residuos y las ventas, con lo que aumenta la posibilidad de que la regresión (11.7.3) presente heteroscedasticidad.

**Prueba de White**

$$\widehat{\hat{u}_i^2} = -46\,746\,325 + 578 \text{ Ventas}_i + 0.000846 \text{ Ventas}_i^2$$
$$\text{ee} = (112\,224\,348) \quad (1\,308) \quad (0.003171) \tag{11.7.6}$$
$$t = (-0.42) \quad (0.44) \quad (0.27)$$
$$R^2 = 0.435$$

Con el valor $R^2$ y $n = 14$, se obtiene $nR^2 = 6.090$. Con la hipótesis nula de inexistencia de heteroscedasticidad, esto debe seguir una distribución ji cuadrada con 2 gl [porque hay dos regresoras en (11.7.6)]. El valor $p$, que resulta de obtener un valor ji cuadrado de 6.090 o mayor, es de casi 0.0476. Como se trata de un valor bajo, la prueba de White también indica heteroscedasticidad.

En resumen, con base en las gráficas de los residuos y las pruebas de Park, Glejser y White, parece que en la regresión de IyD (11.7.3) existe heteroscedasticidad. Como la verdadera varianza del error se desconoce, no podemos utilizar el método de los mínimos cuadrados ponderados para obtener los errores estándares corregidos por heteroscedasticidad y los valores *t*. Por consiguiente, con base en los datos disponibles, tendríamos que hacer conjeturas respecto de la naturaleza de la varianza de error.

Para concluir el ejemplo, a continuación presentamos los errores estándar de White consistentes con heteroscedasticidad, como se analizaron en la sección 11.6:

$$\begin{aligned} \widehat{\text{IyD}}_i &= 1\,337.87 \quad + \; 0.0437\ \text{Ventas}_i \\ \text{ee} &= (4\,892.447) \quad (0.0411) \\ t &= \quad (0.27) \qquad (1.06) \qquad\qquad r^2 = 0.172 \end{aligned} \tag{11.7.7}$$

Al comparar la ecuación (11.7.7) con la ecuación (11.7.3) (la segunda sin corrección por heteroscedasticidad), observamos que, a pesar de que no cambiaron los parámetros estimados (como era de esperarse), el error estándar del coeficiente del intercepto disminuyó un poco y el error estándar del coeficiente de la pendiente se incrementó un poco. Pero recuerde que el procedimiento de White es estrictamente para muestras grandes, en tanto que aquí sólo tenemos 14 observaciones.

**EJEMPLO 11.11**

La tabla 11.6 del sitio Web del libro contiene datos sobre salarios y otros aspectos relacionados correspondientes a 94 distritos escolares en el noroeste de Ohio. En principio, se estimó la siguiente regresión con base en estos datos:

$$\ln(\text{Salario})_i = \beta_1 + \beta_2 \ln(\text{Ingresofam}) + \beta_3 \ln(\text{Valorinm}) + u_i$$

Donde Salario = salario promedio de los maestros ($), Ingresofam = ingreso familiar promedio en el distrito ($) y Valorinm = valor promedio de los inmuebles en el distrito ($).

Como se trata de un modelo de doble logaritmo, todos los coeficientes de las pendientes son elasticidades. Con base en las diversas pruebas de heteroscedasticidad analizadas en el texto, se concluyó que el modelo anterior tenía heteroscedasticidad. Por tanto, obtuvimos los errores estándar robustos (de White). La siguiente tabla presenta los resultados de la regresión precedente con y sin errores estándar robustos.

| Variable | Coeficiente | ee MCO | ee robusto |
|---|---|---|---|
| Intercepto | 7.0198 | 0.8053 | 0.7721 |
| | | (8.7171) | (9.0908) |
| ln(ingresofam) | 0.2575 | 0.0799 | 0.1009 |
| | | (3.2230) | (2.5516) |
| ln(valorinm) | 0.0704 | 0.0207 | 0.0460 |
| | | (3.3976) | (1.5311) |
| $R^2$ | 0.2198 | | |

*Nota:* Las cifras entre paréntesis son razones *t* estimadas.

Aunque los valores de los coeficientes y $R^2$ son iguales con el método de MCO o con el de White, los errores estándar cambiaron; el cambio más radical es el del error estándar del coeficiente ln(valorinm). El método habitual de MCO indicaría que el coeficiente estimado de esta variable es muy significativo estadísticamente, mientras que el error estándar robusto de White indica que este coeficiente no es significativo ni siquiera en el nivel de 10%. Lo importante de este ejemplo es que, si existe heteroscedasticidad, debe tomarse en cuenta para estimar el modelo.

## 11.8 Advertencia respecto de una reacción exagerada ante la heteroscedasticidad

En el ejemplo sobre IyD analizado en la sección anterior vimos que, cuando utilizamos la transformación raíz cuadrada para corregir la heteroscedasticidad en el modelo original (11.7.3), el error estándar del coeficiente de la pendiente disminuyó y su valor $t$ aumentó. ¿Este cambio es tan importante como para causar preocupación en la práctica? En otras palabras, ¿cuándo debe preocupar el problema de la heteroscedasticidad? Como sostiene un autor: "la heteroscedasticidad jamás ha sido una razón para desechar un modelo que de otra forma sería adecuado".[43]

Aquí vale la pena recordar la advertencia de John Fox:

> . . .vale la pena corregir una varianza del error desigual sólo cuando el problema es grave.
>
> El impacto de una varianza del error no constante sobre la eficacia de un estimador de mínimos cuadrados ordinarios y sobre la validez de la inferencia de mínimos cuadrados depende de diversos factores, como el tamaño de la muestra, el grado de variación de $\sigma_i^2$, la configuración de los valores $X$ (es decir, la regresora) y de la relación entre la varianza del error y las $X$. Por consiguiente, no es posible llegar a conclusiones generales respecto del daño producido por la heteroscedasticidad.[44]

De regreso al modelo (11.3.1), vimos ya que la varianza del estimador de pendiente, var $(\hat{\beta}_2)$, está dada por la fórmula usual mostrada en (11.2.3). Según MCG, la varianza del estimador de pendiente, var $(\hat{\beta}_2^*)$, está dada por (11.3.9). Sabemos que esta última es más eficaz que la primera. Pero, ¿qué tan grande debe ser la primer varianza (es decir, la de MCO) en relación con la varianza de MCG antes de que cause preocupación? Como regla práctica, Fox sugiere que el problema empiece a preocupar "...cuando la varianza del error más grande sea mayor que 10 veces la más pequeña".[45] Por consiguiente, al revisar los resultados de las simulaciones Monte Carlo de Davidson y MacKinnon presentadas antes, considere el valor de $\alpha = 2$. La varianza del $\beta_2$ estimado es 0.04 con MCO, y 0.012 con MCG, por lo que la razón de la primera respecto de la segunda es de casi 3.33.[46] De acuerdo con la regla de Fox, en este caso la gravedad de la heteroscedasticidad quizá no sea lo bastante grande para provocar preocupación.

También recuerde que, a pesar de la heteroscedasticidad, los estimadores de MCO (en condiciones generales) son lineales e insesgados, y están asintótica y normalmente distribuidos (es decir, en muestras grandes).

Como veremos cuando analicemos otra violación a los supuestos del modelo clásico de regresión lineal, la advertencia de esta sección resulta apropiada como regla general. Si hace caso omiso de lo anterior, puede cometer errores.

## Resumen y conclusiones

1. Un supuesto importante del modelo clásico de regresión lineal es que todas las perturbaciones $u_i$ tienen la misma varianza $\sigma^2$. Si este supuesto no se satisface, hay heteroscedasticidad.
2. La heteroscedasticidad no destruye las propiedades de insesgamiento y consistencia de los estimadores de MCO.
3. Sin embargo, estos estimadores dejan de tener varianza mínima, es decir, de ser eficientes. Por consiguiente, no son MELI.
4. Los estimadores MELI son proporcionados por el método de mínimos cuadrados ponderados, siempre que se conozcan las varianzas heteroscedásticas de error, $\sigma_i^2$.

[43] N. Gregory Mankiw, "A Quick Refresher Course in Macroeconomics", *Journal of Economic Literature,* vol. XXVIII, diciembre de 1990, p. 1648.

[44] John Fox, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage, California, 1997, p. 306.

[45] *Ibid.*, p. 307.

[46] Observe que los errores estándar se elevaron al cuadrado para obtener las varianzas.

5. En presencia de heteroscedasticidad, las varianzas de los estimadores de MCO no se obtienen con las fórmulas usuales de MCO. Sin embargo, si insiste en utilizar las fórmulas habituales de MCO, las pruebas $t$ y $F$ basadas en ellas pueden conducir a grandes desatinos que darán por resultado conclusiones erróneas.
6. Es más fácil documentar las consecuencias de la heteroscedasticidad que detectarlas. Existen diversas pruebas de diagnóstico disponibles, pero no se puede decir con seguridad cuál funcionará en una situación dada.
7. Aunque se sospeche y se detecta la heteroscedasticidad, no es fácil corregir el problema. Si la muestra es grande, se pueden obtener los errores estándar de los estimadores de MCO corregidos por el método de heteroscedasticidad de White y realizar inferencia estadística basada en estos errores estándar.
8. De lo contrario, con base en los residuos de MCO, se pueden hacer conjeturas con ciertos fundamentos acerca del patrón probable de heteroscedasticidad y transformar la información original de manera que en la información transformada no haya heteroscedasticidad.

## EJERCICIOS

### Preguntas

11.1. Establezca si las siguientes afirmaciones son verdaderas, falsas o inciertas y *comente sus razones brevemente*:

*a*) En presencia de heteroscedasticidad, los estimadores de MCO son sesgados e ineficientes.

*b*) Si hay heteroscedasticidad, las pruebas convencionales $t$ y $F$ son inválidas.

*c*) En presencia de heteroscedasticidad, el método de MCO habitual siempre sobreestima los errores estándar de los estimadores.

*d*) Si los residuales estimados mediante una regresión por MCO exhiben un patrón sistemático, significa que hay heteroscedasticidad en los datos.

*e*) No hay una prueba general de heteroscedasticidad que no esté basada en algún supuesto acerca de cuál variable está correlacionada con el término de error.

*f*) Si el modelo de regresión está mal especificado (por ejemplo, si se omitió una variable importante), los residuos de MCO mostrarán un patrón claramente distinguible.

*g*) Si una regresora con varianza no constante se omite (incorrectamente) de un modelo, los residuos (MCO) serán heteroscedásticos.

11.2. En una regresión de salarios promedio ($W$, \$) sobre el número de empleados ($N$) de una muestra aleatoria de 30 empresas se obtuvieron los siguientes resultados:*

$$\widehat{W} = 7.5 + 0.009N \qquad \qquad (1)$$
$$t = \text{n.a.} \quad (16.10) \qquad R^2 = 0.90$$
$$\widehat{W}/N = 0.008 + 7.8(1/N) \qquad \qquad (2)$$
$$t = (14.43) \quad (76.58) \qquad R^2 = 0.99$$

*a*) ¿Cómo interpreta las dos regresiones?

*b*) ¿Qué supone el autor al pasar de la ecuación (1) a la (2)? ¿Le preocupaba la heteroscedasticidad? ¿Cómo sabe?

*c*) ¿Puede relacionar las pendientes y los interceptos de los dos modelos?

*d*) ¿Puede comparar los valores de $R^2$ de los dos modelos? ¿Por qué?

* Véase Dominick Salvatore, *Managerial Economics*, McGraw-Hill, Nueva York, 1989, p. 157.

11.3. *a*) ¿Puede estimar los parámetros de los modelos

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

mediante el método de mínimos cuadrados ordinarios? ¿Por qué?

*b*) Si la respuesta es negativa, ¿puede sugerir un método informal o formal de estimación de los parámetros de tales modelos? (Véase el capítulo 14.)

11.4. Aunque los modelos logarítmicos como el de la ecuación (11.6.12) a menudo reducen la heteroscedasticidad, se debe prestar cuidadosa atención a las propiedades del término de perturbación de estos modelos. Por ejemplo, el modelo

$$Y_i = \beta_1 X_i^{\beta_2} u_i \tag{1}$$

puede escribirse como

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \ln u_i \tag{2}$$

*a*) Si ln $u_i$ tiene valor esperado cero, ¿cuál debe ser la distribución de $u_i$?

*b*) Si $E(u_i) = 1$, ¿será $E(\ln u_i) = 0$? ¿Por qué?

*c*) Si $E(\ln u_i)$ es diferente de cero, ¿qué puede hacer para volverlo cero?

11.5. Muestre que $\beta_2^*$ de (11.3.8) también se expresa como

$$\beta_2^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{2*}}$$

y var ($\beta_2^*$) dada en (11.3.9) también se expresa como

$$\text{var}\,(\beta_2^*) = \frac{1}{\sum w_i x_i^{2*}}$$

donde $y_i^* = Y_i - \bar{Y}^*$ y $x_i^* = X_i - \bar{X}^*$ representan las desviaciones en relación con las medias ponderadas $\bar{Y}^*$ y $\bar{X}^*$ definidas como

$$\bar{Y}^* = \sum w_i Y_i \Big/ \sum w_i$$

$$\bar{X}^* = \sum w_i X_i \Big/ \sum w_i$$

11.6. Con propósitos pedagógicos, Hanushek y Jackson estiman el siguiente modelo:

$$C_t = \beta_1 + \beta_2 \text{PNB}_t + \beta_3 D_t + u_i \tag{1}$$

donde $C_t$ = gasto agregado de consumo privado en el año $t$, $\text{PNB}_t$ = producto nacional bruto en el año $t$ y $D_t$ = gastos de defensa nacional en el año $t$, con el objetivo de estudiar el efecto de los gastos de defensa sobre otros gastos en la economía.

Los autores postulan que $\sigma_t^2 = \sigma^2(\text{PNB}_t)^2$, luego transforman (1) y estiman

$$C_t/\text{PIB}_t = \beta_1\,(1/\text{PIB}_t) + \beta_2 + \beta_3\,(D_t/\text{PIB}_t) + u_t/\text{PIB}_t \tag{2}$$

Los resultados empíricos basados en la información de 1946 a 1975 fueron los siguientes (errores estándar entre paréntesis):*

$$\hat{C}_t = \underset{(2.73)}{26.19} + \underset{(0.0060)}{0.6248}\,\text{PNB}_t - \underset{(0.0736)}{0.4398}\,D_t \qquad R^2 = 0.999$$

$$\widehat{C_t/\text{PNB}_t} = \underset{(2.22)}{25.92}\,(1/\text{PNB}_t) + \underset{(0.0068)}{0.6246} - \underset{(0.0597)}{0.4315}\,(D_t/\text{PNB}_t) \qquad R^2 = 0.875$$

*a*) ¿Qué supuesto hacen los autores sobre la naturaleza de la heteroscedasticidad? ¿Puede justificarlo?

*b*) Compare los resultados de las dos regresiones. ¿La transformación del modelo original mejora los resultados, es decir, reduce los errores estándar estimados? ¿Por qué?

*c*) ¿Puede comparar los dos valores de $R^2$? ¿Por qué? (*Sugerencia:* Examine las variables dependientes.)

11.7. Consulte las regresiones estimadas (11.6.2) y (11.6.3). Los resultados de la regresión son muy similares. ¿A qué se debe esta conclusión?

11.8. Pruebe que si $w_i = w$, una constante, para cada $i$, $\beta_2^*$ y $\hat{\beta}_2$ son idénticos, así como sus varianzas.

11.9. Consulte las fórmulas (11.2.2) y (11.2.3), y suponga que

$$\sigma_i^2 = \sigma^2 k_i$$

donde $\sigma^2$ es una constante y $k_i$ son ponderaciones *conocidas*, no necesariamente todas iguales.

Con este supuesto, muestre que la varianza dada en (11.2.2) se expresa como

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 k_i}{\sum x_i^2}$$

El primer término del miembro derecho es la fórmula de la varianza dada en (11.2.3), es decir, var $(\beta_2^*)$ con homoscedasticidad. ¿Qué puede decir sobre la naturaleza de la relación entre var $(\hat{\beta}_2)$ con heteroscedasticidad y con homoscedasticidad? (*Sugerencia:* Examine, en la fórmula anterior, el segundo término del miembro derecho.) ¿Puede derivar alguna conclusión general sobre las relaciones entre (11.2.2) y (11.2.3)?

11.10. En el modelo

$$Y_i = \beta_2 X_i + u_i \quad (\textit{Nota:}\ \text{No hay intercepto})$$

se le informa que $\text{var}\,(u_i) = \sigma^2 X_i^2$. Demuestre que

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2 \sum X_i^4}{\left(\sum X_i^2\right)^2}$$

* Eric A. Hanushek y John E. Jackson, *Statistical Methods for Social Scientists*, Academic, Nueva York, 1977, p. 160.

## Ejercicios empíricos

11.11. Con la información de la tabla 11.1, efectúe la regresión de la remuneración salarial promedio $Y$ sobre la productividad promedio $X$, y considere el tamaño de la planta laboral como unidad de observación. Interprete sus resultados y vea si están de acuerdo con los presentados en (11.5.3).

*a*) De la regresión anterior, obtenga los residuos $\hat{u}_i$.

*b*) Según la prueba de Park, efectúe la regresión de ln $\hat{u}_i^2$ sobre ln $X_i$ y verifique la regresión (11.5.4).

*c*) Según el método de Glejser, efectúe la regresión de $|\hat{u}_i|$ sobre $X_i$ y luego la regresión de $|\hat{u}_i|$ sobre $\sqrt{X_i}$. Comente sus resultados.

*d*) Encuentre la correlación de orden entre $|\hat{u}_i|$ y $X_i$, y comente sobre la naturaleza de la heteroscedasticidad presente en los datos, si existe.

11.12. La tabla 11.6 presenta información sobre la razón ventas/efectivo en las industrias manufactureras de Estados Unidos, clasificadas por tamaño de activos del establecimiento de 1971-I a 1973-IV. (Información trimestral.) La razón ventas/efectivo puede considerarse una medida de la velocidad del ingreso en el sector empresarial, es decir, el número de veces que circula un dólar.

*a*) Por cada tamaño de activos, calcule la media y la desviación estándar de la razón ventas/efectivo.

*b*) Grafique el valor de la media frente a la desviación estándar obtenida en *a*), con el tamaño de activos como unidad de observación.

*c*) Con un modelo de regresión apropiado, determine si la desviación estándar de la razón se incrementa con el valor de la media. De no ser así, ¿cómo interpreta el resultado?

*d*) Si hay una relación estadísticamente significativa entre los dos, ¿cómo transformaría la información de manera que no haya heteroscedasticidad?

11.13. *Prueba de homogeneidad de varianza de Bartlett.** Suponga que hay $k$ varianzas muestrales independientes $s_1^2, s_2^2, \ldots, s_k^2$ con $f_1, f_2, \ldots, f_k$ gl, cada una proveniente de poblaciones normalmente distribuidas con media $\mu$ y varianza $\sigma_i^2$. Suponga además que deseamos probar la hipótesis nula $H_0$: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$; es decir, cada varianza muestral es una estimación de la misma varianza poblacional $\sigma^2$.

Si la hipótesis nula es verdadera, entonces

$$s^2 = \frac{\sum_{i=1}^{k} f_i s_i^2}{\sum f_i} = \frac{\sum f_i s_i^2}{f}$$

**TABLA 11.6**
**Tamaño de activos (millones de dólares)**

Fuente: *Quarterly Financial Report for Manufacturing Corporations*, Federal Trade Commission y Securities and Exchange Commission, gobierno de Estados Unidos, variables sobre diversos temas (calculadas).

| Año y trimestre | 1-10 | 10-25 | 25-50 | 50-100 | 100-250 | 250-1 000 | 1 000 + |
|---|---|---|---|---|---|---|---|
| 1971-I | 6.696 | 6.929 | 6.858 | 6.966 | 7.819 | 7.557 | 7.860 |
| -II | 6.826 | 7.311 | 7.299 | 7.081 | 7.907 | 7.685 | 7.351 |
| -III | 6.338 | 7.035 | 7.082 | 7.145 | 7.691 | 7.309 | 7.088 |
| -IV | 6.272 | 6.265 | 6.874 | 6.485 | 6.778 | 7.120 | 6.765 |
| 1972-I | 6.692 | 6.236 | 7.101 | 7.060 | 7.104 | 7.584 | 6.717 |
| -II | 6.818 | 7.010 | 7.719 | 7.009 | 8.064 | 7.457 | 7.280 |
| -III | 6.783 | 6.934 | 7.182 | 6.923 | 7.784 | 7.142 | 6.619 |
| -IV | 6.779 | 6.988 | 6.531 | 7.146 | 7.279 | 6.928 | 6.919 |
| 1973-I | 7.291 | 7.428 | 7.272 | 7.571 | 7.583 | 7.053 | 6.630 |
| -II | 7.766 | 9.071 | 7.818 | 8.692 | 8.608 | 7.571 | 6.805 |
| -III | 7.733 | 8.357 | 8.090 | 8.357 | 7.680 | 7.654 | 6.772 |
| -IV | 8.316 | 7.621 | 7.766 | 7.867 | 7.666 | 7.380 | 7.072 |

* Véase "Properties of Sufficiency and Statistical Tests", *Proceedings of the Royal Society of London A*, vol. 160, 1937, p. 268.

constituye una estimación de la estimación común (agrupada) de la varianza poblacional $\sigma^2$, donde $f_i = (n_i - 1)$, con $n_i$ como el número de observaciones en el $i$-ésimo grupo y donde $f = \sum_{i=1}^{k} f_i$.

Bartlett demostró que la hipótesis nula se prueba por la razón $A/B$, distribuida aproximadamente como la distribución $\chi^2$ con $k - 1$ gl, donde

$$A = f \ln s^2 - \sum \left( f_i \ln s_i^2 \right)$$

y

$$B = 1 + \frac{1}{3(k-1)} \left[ \sum \left( \frac{1}{f_i} \right) - \frac{1}{f} \right]$$

Aplique la prueba de Bartlett a los datos de la tabla 11.1 y verifique que no se puede rechazar la hipótesis de que las varianzas poblacionales de la remuneración salarial son las mismas para cada tamaño de la planta laboral del establecimiento, en el nivel de significancia de 5%.

*Nota:* $f_i$, los gl para cada varianza muestral, es 9, pues $n_i$ para cada muestra (es decir, clase de empleados) es 10.

11.14. Considere el siguiente modelo de regresión a través del origen:

$$Y_i = \beta X_i + u_i, \quad \text{para } i = 1, 2$$

Se tiene que $u_1 \sim N(0, \sigma^2)$ y $u_2 \sim N(0, 2\sigma^2)$, y que son estadísticamente independientes. Si $X_1 = +1$ y $X_2 = -1$, obtenga la estimación por mínimos cuadrados *ponderados* (MCP) de $\beta$ y su varianza. Si en esta situación supuso de manera incorrecta que la dos varianzas de los errores son iguales (por ejemplo, iguales a $\sigma^2$), ¿cuál sería el estimador de MCO de $\beta$?, ¿y su varianza? Compare estas estimaciones con las obtenidas por el método de MCP. ¿Qué conclusión general deduce?*

11.15. La tabla 11.7 proporciona datos sobre 81 automóviles respecto de su MPG (millas promedio por galón), CF (caballos de fuerza de su motor), VOL (pies cúbicos de su cabina), VM (velocidad máxima en millas por hora) y su PS (peso del vehículo en cientos de lb).

*a*) Considere el siguiente modelo:

$$\text{MPG}_i = \beta_1 + \beta_2 \text{VM}_i + \beta_3 \text{CF}_i + \beta_4 \text{PS}_i + u_i$$

Estime los parámetros de este modelo e interprete los resultados. Desde el punto de vista económico, ¿tiene sentido?

*b*) ¿Esperaría que la varianza del error en el modelo anterior sea heteroscedástica? ¿Por qué?

*c*) Con la prueba de White determine si la varianza de error es heteroscedástica.

*d*) Obtenga los errores estándar de White consistentes con la heteroscedasticidad, así como los valores $t$, y compare los resultados con los obtenidos mediante MCO.

*e*) Si se establece heteroscedasticidad, ¿cómo puede transformar los datos de manera que en los datos transformados la varianza del error sea homoscedástica? Muestre los cálculos necesarios.

11.16. *Gasto alimentario en India.* En la tabla 2.8 se proporcionaron datos sobre el gasto en alimentos y el gasto total de 55 familias de India.

*a*) Haga la regresión del gasto alimentario sobre el gasto total y examine los residuos obtenidos en dicha regresión.

*b*) Grafique los residuos obtenidos en el inciso *a*) contra el gasto total y verifique si existe algún patrón sistemático.

---

* Adaptado de F.A.F. Seber, *Linear Regression Analysis*, John Wiley & Sons, Nueva York, 1977, p. 64.

**TABLA 11.7 Datos de millajes de automóviles de pasajeros**

| Observación | MPG | VM | CF | VOL | PS | Observación | MPG | VM | CF | VOL | PS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 65.4 | 96 | 49 | 89 | 17.5 | 42 | 32.2 | 106 | 95 | 106 | 30.0 |
| 2 | 56.0 | 97 | 55 | 92 | 20.0 | 43 | 32.2 | 109 | 102 | 92 | 30.0 |
| 3 | 55.9 | 97 | 55 | 92 | 20.0 | 44 | 32.2 | 106 | 95 | 88 | 30.0 |
| 4 | 49.0 | 105 | 70 | 92 | 20.0 | 45 | 31.5 | 105 | 93 | 102 | 30.0 |
| 5 | 46.5 | 96 | 53 | 92 | 20.0 | 46 | 31.5 | 108 | 100 | 99 | 30.0 |
| 6 | 46.2 | 105 | 70 | 89 | 20.0 | 47 | 31.4 | 108 | 100 | 111 | 30.0 |
| 7 | 45.4 | 97 | 55 | 92 | 20.0 | 48 | 31.4 | 107 | 98 | 103 | 30.0 |
| 8 | 59.2 | 98 | 62 | 50 | 22.5 | 49 | 31.2 | 120 | 130 | 86 | 30.0 |
| 9 | 53.3 | 98 | 62 | 50 | 22.5 | 50 | 33.7 | 109 | 115 | 101 | 35.0 |
| 10 | 43.4 | 107 | 80 | 94 | 22.5 | 51 | 32.6 | 109 | 115 | 101 | 35.0 |
| 11 | 41.1 | 103 | 73 | 89 | 22.5 | 52 | 31.3 | 109 | 115 | 101 | 35.0 |
| 12 | 40.9 | 113 | 92 | 50 | 22.5 | 53 | 31.3 | 109 | 115 | 124 | 35.0 |
| 13 | 40.9 | 113 | 92 | 99 | 22.5 | 54 | 30.4 | 133 | 180 | 113 | 35.0 |
| 14 | 40.4 | 103 | 73 | 89 | 22.5 | 55 | 28.9 | 125 | 160 | 113 | 35.0 |
| 15 | 39.6 | 100 | 66 | 89 | 22.5 | 56 | 28.0 | 115 | 130 | 124 | 35.0 |
| 16 | 39.3 | 103 | 73 | 89 | 22.5 | 57 | 28.0 | 102 | 96 | 92 | 35.0 |
| 17 | 38.9 | 106 | 78 | 91 | 22.5 | 58 | 28.0 | 109 | 115 | 101 | 35.0 |
| 18 | 38.8 | 113 | 92 | 50 | 22.5 | 59 | 28.0 | 104 | 100 | 94 | 35.0 |
| 19 | 38.2 | 106 | 78 | 91 | 22.5 | 60 | 28.0 | 105 | 100 | 115 | 35.0 |
| 20 | 42.2 | 109 | 90 | 103 | 25.0 | 61 | 27.7 | 120 | 145 | 111 | 35.0 |
| 21 | 40.9 | 110 | 92 | 99 | 25.0 | 62 | 25.6 | 107 | 120 | 116 | 40.0 |
| 22 | 40.7 | 101 | 74 | 107 | 25.0 | 63 | 25.3 | 114 | 140 | 131 | 40.0 |
| 23 | 40.0 | 111 | 95 | 101 | 25.0 | 64 | 23.9 | 114 | 140 | 123 | 40.0 |
| 24 | 39.3 | 105 | 81 | 96 | 25.0 | 65 | 23.6 | 117 | 150 | 121 | 40.0 |
| 25 | 38.8 | 111 | 95 | 89 | 25.0 | 66 | 23.6 | 122 | 165 | 50 | 40.0 |
| 26 | 38.4 | 110 | 92 | 50 | 25.0 | 67 | 23.6 | 122 | 165 | 114 | 40.0 |
| 27 | 38.4 | 110 | 92 | 117 | 25.0 | 68 | 23.6 | 122 | 165 | 127 | 40.0 |
| 28 | 38.4 | 110 | 92 | 99 | 25.0 | 69 | 23.6 | 122 | 165 | 123 | 40.0 |
| 29 | 46.9 | 90 | 52 | 104 | 27.5 | 70 | 23.5 | 148 | 245 | 112 | 40.0 |
| 30 | 36.3 | 112 | 103 | 107 | 27.5 | 71 | 23.4 | 160 | 280 | 50 | 40.0 |
| 31 | 36.1 | 103 | 84 | 114 | 27.5 | 72 | 23.4 | 121 | 162 | 135 | 40.0 |
| 32 | 36.1 | 103 | 84 | 101 | 27.5 | 73 | 23.1 | 121 | 162 | 132 | 40.0 |
| 33 | 35.4 | 111 | 102 | 97 | 27.5 | 74 | 22.9 | 110 | 140 | 160 | 45.0 |
| 34 | 35.3 | 111 | 102 | 113 | 27.5 | 75 | 22.9 | 110 | 140 | 129 | 45.0 |
| 35 | 35.1 | 102 | 81 | 101 | 27.5 | 76 | 19.5 | 121 | 175 | 129 | 45.0 |
| 36 | 35.1 | 106 | 90 | 98 | 27.5 | 77 | 18.1 | 165 | 322 | 50 | 45.0 |
| 37 | 35.0 | 106 | 90 | 88 | 27.5 | 78 | 17.2 | 140 | 238 | 115 | 45.0 |
| 38 | 33.2 | 109 | 102 | 86 | 30.0 | 79 | 17.0 | 147 | 263 | 50 | 45.0 |
| 39 | 32.9 | 109 | 102 | 86 | 30.0 | 80 | 16.7 | 157 | 295 | 119 | 45.0 |
| 40 | 32.3 | 120 | 130 | 92 | 30.0 | 81 | 13.2 | 130 | 236 | 107 | 55.0 |
| 41 | 32.2 | 106 | 95 | 113 | 30.0 | | | | | | |

*Notas:*

VOL = pies cúbicos del espacio de cabina
CF = caballos de fuerza del motor
MPG = millas promedio por galón
VM = velocidad máxima, millas por hora
PS = peso del vehículo, cientos de libras
Observación = número de observaciones del automóvil (el nombre de los vehículos no se proporciona).

Fuente: U.S. Environmental Protection Agency, 1991, Informe EPA/AA/CTAB/91-02.

*c*) Si la gráfica del inciso *b*) sugiere heteroscedasticidad, aplique las pruebas de Park, Glejser y White para determinar si la sensación respecto de la heteroscedasticidad observada en *b*) se sustenta con estas pruebas.

*d*) Obtenga los errores estándar de White consistentes con la heteroscedasticidad y compárelos con los errores estándar de MCO. Decida si vale la pena corregir este ejemplo a causa de la heteroscedasticidad.

11.17. Repita el ejercicio 11.16, pero en esta ocasión efectúe la regresión del logaritmo del gasto alimentario sobre el logaritmo del gasto total. Si observa heteroscedasticidad en el modelo lineal del ejercicio 11.16 pero no en el modelo log-lineal, ¿a qué conclusión llega? Muestre todos los cálculos necesarios.

11.18. *Un atajo de la prueba de White.* Como mencionamos en el texto, la prueba de White consume grados de libertad si existen varias regresoras y se introducen todas las regresoras, sus términos cuadrados y sus productos cruzados. Por consiguiente, en vez de estimar las regresiones como la (11.5.22), ¿por qué no simplemente efectúa la siguiente regresión?

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 \hat{Y}_i + \alpha_2 \hat{Y}_i^2 + \nu_i$$

donde $\hat{Y}_i$ son los valores estimados $Y$ (es decir, la regresada) de cualquier modelo que se calcule. Después de todo, $\hat{Y}_i$ es tan sólo el promedio ponderado de las regresoras, donde los coeficientes estimados de la regresión sirven como ponderaciones.

Obtenga el valor $R^2$ de la regresión anterior y utilice (11.5.22) para probar la hipótesis de que no existe heteroscedasticidad.

Aplique la prueba anterior al ejemplo de gasto alimentario del ejercicio 11.6.

11.19. Reconsidere el ejemplo sobre IyD de la sección 11.7. Repita ese ejemplo con las ganancias como la regresora. *A priori,* ¿esperaría que los resultados fuesen diferentes de los que utilizan las ventas como regresoras?, ¿por qué?

11.20. La tabla 11.8 proporciona datos sobre la mediana de los salarios de catedráticos en estadística que laboraron en centros universitarios de investigación de Estados Unidos durante el año académico 2007.

*a*) Grafique la mediana de los salarios respecto de los rangos de años (como medida de los años de experiencia). Para propósitos de la gráfica, suponga que la mediana de los salarios está referida al punto medio del rango de años correpondiente. Por consiguiente, el salario de $124 578 del rango 4-5 está referido a 4.5 años del rango correspondiente, y así sucesivamente. Para el último grupo, suponga que el rango es 31-33.

*b*) Considere los siguientes modelos de regresión:

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \quad (1)$$

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \nu_i \quad (2)$$

**TABLA 11.8**
**Mediana de los salarios de los catedráticos en estadística, 2007**

Fuente: American Statistical Association, "2007 Salary Report".

| Rango de años | Cuenta | Mediana |
|---|---|---|
| 0 a 1 | 40 | $101 478 |
| 2 a 3 | 24 | 102 400 |
| 4 a 5 | 35 | 124 578 |
| 6 a 7 | 34 | 122 850 |
| 8 a 9 | 33 | 116 900 |
| 10 a 14 | 73 | 119 465 |
| 15 a 19 | 69 | 114 900 |
| 20 a 24 | 54 | 129 072 |
| 25 a 30 | 44 | 131 704 |
| 31 o más | 25 | 143 000 |

donde $Y$ = mediana del salario, $X$ = año en el rango (medido como el punto medio del intervalo), y $u$ y $v$ son los términos de error. ¿Puede justificar por qué el modelo (2) sería preferible al modelo (1)? A partir de estos datos, estime los modelos.

*c*) Si observa heteroscedasticidad en el modelo (1) pero no en el modelo (2), ¿a qué conclusiones llega? Muestre los cálculos necesarios.

*d*) Si observa heteroscedasticidad en el modelo (2), ¿cómo puede transformar los datos de manera que en el modelo transformado no existiera heteroscedasticidad?

11.21. Tiene la siguiente información:

$$\text{SCR}_1 \text{ basado en las primeras 30 observaciones} = 55, \text{gl} = 25$$
$$\text{SCR}_2 \text{ basado en las últimas 30 observaciones} = 140, \text{gl} = 25$$

Realice la prueba de heteroscedasticidad de Goldfeld-Quandt en el nivel de significancia de 5%.

11.22. La tabla 11.9 presenta información acerca de los precios de acciones ($Y$) y los precios al consumidor ($X$) expresados en cambios porcentuales anuales para un corte transversal de 20 países.

*a*) Grafique los datos en un diagrama de dispersión.

*b*) Efectúe la regresión de $Y$ sobre $X$ y examine los residuos de esta regresión. ¿Qué observa?

*c*) Como los datos de Chile parecen atípicos, repita la regresión en *b*) sin la información sobre Chile. Ahora examine los residuos de esta regresión. ¿Qué observa?

*d*) Si, con base en los resultados de *b*), concluye que hubo heteroscedasticidad en la varianza del error, pero con base en los resultados de *c*) modifica este resultado, ¿qué conclusiones generales obtiene?

**TABLA 11.9**
**Precios de acciones y precios al consumidor, periodo posterior a la Segunda Guerra Mundial (hasta 1969)**

Fuente: Phillip Cagan, *Common Stock Values and Inflation: The Historical Record of Many Countries*, National Bureau of Economic Research, suplemento, marzo de 1974, tabla 1, p. 4.

| | Tasa de crecimiento, % anual | |
|---|---|---|
| **País** | **Precio de acciones, $Y$** | **Precio al consumidor, $X$** |
| 1. Australia | 5.0 | 4.3 |
| 2. Austria | 11.1 | 4.6 |
| 3. Bélgica | 3.2 | 2.4 |
| 4. Canadá | 7.9 | 2.4 |
| 5. Chile | 25.5 | 26.4 |
| 6. Dinamarca | 3.8 | 4.2 |
| 7. Finlandia | 11.1 | 5.5 |
| 8. Francia | 9.9 | 4.7 |
| 9. Alemania | 13.3 | 2.2 |
| 10. India | 1.5 | 4.0 |
| 11. Irlanda | 6.4 | 4.0 |
| 12. Israel | 8.9 | 8.4 |
| 13. Italia | 8.1 | 3.3 |
| 14. Japón | 13.5 | 4.7 |
| 15. México | 4.7 | 5.2 |
| 16. Países Bajos | 7.5 | 3.6 |
| 17. Nueva Zelanda | 4.7 | 3.6 |
| 18. Suecia | 8.0 | 4.0 |
| 19. Reino Unido | 7.5 | 3.9 |
| 20. Estados Unidos | 9.0 | 2.1 |

11.23 La tabla 11.10 del sitio Web presenta datos sobre salarios y otros aspectos relacionados de 447 ejecutivos de 500 empresas de *Fortune*. Los datos incluyen salario = salario y gratificaciones de 1999; remtot = remuneración total del director general en 1999; antigüedad = número de años como director general (0 si tiene menos de 6 meses en el cargo); edad = edad del director general; ventas = total de los ingresos por ventas de la empresa en 1998; utilidades = utilidades de la empresa en 1998; y activos = total de activos de la empresa en 1998.

*a*) Estime la siguiente regresión con base en estos datos y obtenga el estadístico de Breusch-Pagan-Godfrey para ver si hay heteroscedasticidad:

$$\text{salario}_i = \beta_1 + \beta_2\text{antigüedad}_i + \beta_3\text{edad}_i + \beta_4\text{ventas}_i + \beta_5\text{utilidades}_i + \beta_6\text{activos}_i + u_i$$

¿Parece existir un problema de heteroscedasticidad?

*b*) Ahora cree un segundo modelo con ln(salario) como variable dependiente. ¿Observa alguna mejora en la heteroscedasticidad?

*c*) Cree diagramas de dispersión del salario sobre cada variable independiente. ¿Puede discernir qué variables contribuyen al problema? ¿Qué propondría ahora para resolverlo? ¿Cuál es el modelo final?

# Apéndice 11A

## 11A.1 Prueba de la ecuación (11.2.2)

Del apéndice 3A, sección 3A.3, tenemos

$$\begin{aligned}\text{var}(\hat\beta_2) &= E\left(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2 + 2 \text{ términos de productos cruzados}\right)\\ &= E\left(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2\right)\end{aligned}$$

pues los valores esperados de los términos de productos cruzados son cero debido al supuesto de no correlación serial,

$$\text{var}(\hat\beta_2) = k_1^2E\left(u_1^2\right) + k_2^2E\left(u_2^2\right) + \cdots + k_n^2E\left(u_n^2\right)$$

porque se conocen las $k_i$. (¿Por qué?)

$$\text{var}(\hat\beta_2) = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2$$

pues $E(u_i^2) = \sigma_i^2$.

$$\begin{aligned}\text{var}(\hat\beta_2) &= \sum k_i^2\sigma_i^2\\ &= \sum\left[\left(\frac{x_i}{\sum x_i^2}\right)^2\sigma_i^2\right] \qquad \text{pues } k_i = \frac{x_i}{\sum x_i^2} \qquad (11.2.2)\\ &= \frac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2}\end{aligned}$$

## 11A.2 Método de mínimos cuadrados ponderados

Para ilustrar el método utilizamos el modelo de dos variables $Y_i = \beta_1 + \beta_2X_i + u_i$. El método de mínimos cuadrados no ponderado reduce

$$\sum\hat u_i^2 = \sum(Y_i - \hat\beta_1 - \hat\beta_2X_i)^2 \qquad (1)$$

para obtener las estimaciones, mientras que el método de mínimos cuadrados ponderados reduce la suma de cuadrados ponderada de los residuales:

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \tag{2}$$

donde $\beta_1^*$ y $\beta_2^*$ son los estimadores de mínimos cuadrados ponderados y las ponderaciones $w_i$ son tales que

$$w_i = \frac{1}{\sigma_i^2} \tag{3}$$

es decir, las ponderaciones son inversamente proporcionales a la varianza de $u_i$ o $Y_i$ condicional a las $X_i$ dadas, y queda entendido que $\text{var}(u_i \mid X_i) = \text{var}(Y_i \mid X_i) = \sigma_i^2$.

Al diferenciar (2) respecto de $\hat{\beta}_1^*$ y $\hat{\beta}_2^*$, obtenemos

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_1^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_2^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-X_i)$$

Igualamos a cero las expresiones anteriores y obtenemos las dos siguientes ecuaciones normales:

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \tag{4}$$

$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2 \tag{5}$$

Observe la semejanza entre estas ecuaciones normales y las ecuaciones normales de los mínimos cuadrados no ponderados.

Resolvemos estas ecuaciones simultáneamente y obtenemos

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6}$$

y

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \tag{11.3.8) = (7}$$

La varianza de $\hat{\beta}_2^*$ que aparece en (11.3.9) se obtiene en la forma de la varianza de $\hat{\beta}_2$ que aparece en el apéndice 3A, sección 3A.3.

*Nota:* $\bar{Y}^* = \sum w_i Y_i / \sum w_i$ y $\bar{X}^* = \sum w_i X_i / \sum w_i$. Como se verifica fácilmente, estas medias ponderadas coinciden con las medias usuales o sin ponderar $\bar{Y}$ y $\bar{X}$ cuando $w_i = w$, una constante, para todo $i$.

## 11A.3 Prueba de que $E(\hat{\sigma}^2) \neq \sigma^2$ en presencia de heteroscedasticidad

Considere el modelo de dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{1}$$

donde $\text{var}(u_i) = \sigma_i^2$.

Ahora

$$\begin{aligned}\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} &= \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum [\beta_1 + \beta_2 X_i + u_i - \hat{\beta}_1 - \hat{\beta}_2 X_i]^2}{n-2} \\ &= \frac{\sum [-(\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2) X_i + u_i]^2}{n-2}\end{aligned} \tag{2}$$

Observe que $(\hat{\beta}_1 - \beta_1) =- (\hat{\beta}_2 - \beta_2)\bar{X} + \bar{u}$, y sustituya lo anterior en (2); al tomar las esperanzas en ambos miembros, tenemos:

$$E(\hat{\sigma}^2) = \frac{1}{n-2}\left\{-\sum x_i^2 \operatorname{var}(\hat{\beta}_2) + E\left[\sum (u_i - \bar{u})^2\right]\right\}$$
$$= \frac{1}{n-2}\left[-\frac{\sum x_i^2\sigma_i^2}{\sum x_i^2} + \frac{(n-1)\sum \sigma_i^2}{n}\right] \tag{3}$$

donde se utiliza (11.2.2).

Como puede observar de (3), si existe homoscedasticidad, es decir, $\sigma_i^2 = \sigma^2$ para cada $i$, $E(\hat{\sigma}^2) = \sigma^2$. En consecuencia, el valor esperado de $\hat{\sigma}^2 = \sum \hat{u}^2/(n-2)$ calculado de manera convencional no será igual a la verdadera $\sigma^2$ en presencia de heteroscedasticidad.[1]

## 11A.4 Errores estándar robustos de White

Para dar alguna idea respecto de los errores estándar de White corregidos por heteroscedasticidad, considere el siguiente modelo de regresión de dos variables:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \operatorname{var}(u_i) = \sigma_i^2 \tag{1}$$

Como se muestra en (11.2.2),

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{2}$$

Como $\sigma_i^2$ no son directamente observables, White sugiere utilizar $\hat{u}_i^2$, el residuo al cuadrado para cada $i$, en lugar de $\sigma_i^2$, y calcular var $(\hat{\beta}_2)$ de la siguiente forma:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_i^2\hat{u}_i^2}{\left(\sum x_i^2\right)^2} \tag{3}$$

White demostró que (3) es un estimador consistente de (2); es decir, conforme el tamaño de la muestra crece de manera indefinida, (3) converge a (2).[2]

A propósito, observe que si el paquete de software utilizado no contiene el procedimiento de errores estándar robustos de White, se puede efectuar como se muestra en (3): primero la regresión usual por MCO, después se obtienen los residuos de esta regresión y luego se utiliza la fórmula (3).

El procedimiento de White se generaliza al modelo de regresión de $k$ variables:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{4}$$

La varianza de todo coeficiente de regresión parcial, por ejemplo $\hat{\beta}_j$, se obtiene así:

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sum \hat{w}_{ji}^2\hat{u}_i^2}{\left(\sum \hat{w}_{ji}^2\right)^2} \tag{5}$$

donde $\hat{u}_i$ son los residuos obtenidos de la regresión original (4), y $\hat{w}_j$ son los residuos proporcionados por la regresión (auxiliar) de la regresora $X_j$ sobre las regresoras restantes en (4).

Obvio, es un procedimiento tardado, pues se debe estimar (5) para cada variable $X$. Por supuesto, todo este trabajo se evita si se cuenta con un software estadístico que lo haga de manera rutinaria. En la actualidad, programas como PC-GIVE, *Eviews*, MICROFIT, SHAZAM, STATA y LIMDEP calculan de manera muy sencilla los errores estándar robustos de White con heteroscedasticidad.

---

[1] Se pueden obtener mayores detalles en Jan Kmenta, *Elements of Econometrics*, 2a. ed., Macmillan, Nueva York, 1986, pp. 276-278.

[2] Para mayor precisión, $n$ veces (3) converge en probabilidad a $E[(X_i - \mu_X)^2 u_i^2]/(\sigma_X^2)^2$, que es la probabilidad límite de $n$ veces (2), donde $n$ es el tamaño de la muestra, $\mu_x$ es el valor esperado de $X$ y $\sigma_X^2$ es la varianza (poblacional) de $X$. Para mayores detalles, véase Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, South-Western, 2000, p. 250.

Capítulo **12**

# Autocorrelación: ¿qué pasa si los términos de error están correlacionados?

El lector quizá recuerde que existen tres tipos de datos disponibles para el análisis empírico: 1) transversales, 2) series de tiempo y 3) la combinación de ambos, también conocida como datos agrupados. Al desarrollar el modelo clásico de regresión lineal (MCRL), en la **parte 1**, partimos de varios supuestos, que se analizaron en la sección 7.1. No obstante, observamos que no todos los supuestos se cumplen con cualquier tipo de datos. De hecho, en el capítulo anterior vimos que el supuesto de la homoscedasticidad, o varianza de error igual, no siempre se sustenta en los datos transversales. En otras palabras, los datos transversales a menudo están plagados de problemas de heteroscedasticidad.

Sin embargo, en los estudios transversales, a menudo los datos se recopilan con base en una muestra aleatoria de unidades transversales; como familias (en un análisis de la función consumo) o empresas (en un estudio de análisis sobre la inversión), de modo que no existe razón previa para creer que el término de error que correspondiente a una familia o a una empresa esté correlacionado con el término de error de otra familia o empresa. Si por casualidad se observa dicha correlación en unidades transversales, se conoce como **autocorrelación espacial**; es decir, es una correlación en el espacio más que en el tiempo. Sin embargo, es importante recordar que en el análisis transversal la ordenación de los datos debe tener alguna lógica, o algún interés económico, a fin de que cobre sentido la conclusión de la presencia o ausencia de autocorrelación (espacial).

No obstante, es probable que la situación sea muy distinta si trabajamos con datos de series de tiempo, pues las observaciones en estos datos siguen un ordenamiento natural respecto del tiempo, de modo que es muy posible que las observaciones sucesivas muestren intercorrelaciones, sobre todo si el intervalo entre observaciones sucesivas es muy corto, como un día, una semana o un mes, en lugar de un año. Si observa los índices bursátiles, como el Dow Jones o el S&P 500 en días sucesivos, no es raro que descubra que dichos índices aumentan o disminuyen durante varios días sucesivos. Obvio, en esta clase de situaciones se viola el supuesto del MCRL en cuanto a que **no existe autocorrelación, ni correlación serial** en los términos de error.

En este capítulo examinaremos en forma crítica este supuesto con el fin de responder las siguientes preguntas:

1. ¿Cuál es la naturaleza de la autocorrelación?
2. ¿Cuáles son las consecuencias teóricas y prácticas de la autocorrelación?

3. Como el supuesto de no autocorrelación se relaciona con las perturbaciones no observables $u_t$, ¿cómo saber que hay autocorrelación en una situación dada? Observe que ahora usaremos el subíndice $t$ para destacar que los datos corresponden a series de tiempo.
4. ¿Cómo remediar el problema de la autocorrelación?

En este capítulo el lector encontrará similitudes en muchos aspectos con el anterior sobre heteroscedasticidad, pues, **en presencia tanto de autocorrelación como de heteroscedasticidad, los estimadores de MCO usuales, a pesar de ser lineales, insesgados y tener distribución asintóticamente normal (es decir, en muestras grandes),[1] dejan de tener varianza mínima entre todos los estimadores lineales insesgados. En resumen, no son eficientes en relación con los demás estimadores lineales e insesgados. Dicho de otro modo, es posible que no sean los mejores estimadores lineales insesgados (MELI). Como resultado, las pruebas usuales *t*, *F* y $\chi^2$ pueden no ser válidas.**

## 12.1 Naturaleza del problema

El término **autocorrelación** se define como la "correlación entre miembros de series de observaciones ordenadas en el tiempo [como en datos de series de tiempo] o en el espacio [como en datos de corte transversal]".[2] En el contexto de regresión, el modelo clásico de regresión lineal supone que no existe tal autocorrelación en las perturbaciones $u_i$. Simbólicamente,

$$\text{cov}(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0 \qquad i \neq j \tag{3.2.5}$$

En forma sencilla, el modelo clásico supone que el término de perturbación relacionado con una observación cualquiera no recibe influencia del término de perturbación relacionado con cualquier otra observación. Por ejemplo, si tratamos con información trimestral de series de tiempo, que implica una regresión de la producción sobre los insumos trabajo y capital, y si, por ejemplo, hay una huelga laboral que afecta la producción en un trimestre, no hay razón para pensar que esta interrupción afectará la producción del trimestre siguiente. Es decir, si la producción es inferior en este trimestre, no hay razón para esperar que sea baja en el siguiente. En forma similar, si tratamos con información de corte transversal que implica la regresión del gasto de consumo familiar sobre el ingreso familiar, no esperaremos que el efecto de un incremento en el ingreso de una familia sobre su gasto de consumo incida en el gasto de consumo de otra.

Sin embargo, si existe tal dependencia, hay autocorrelación. Simbólicamente,

$$E(u_i u_j) \neq 0 \qquad i \neq j \tag{12.1.1}$$

En esta situación, la interrupción ocasionada por una huelga este trimestre puede afectar muy fácilmente la producción del siguiente trimestre, o los incrementos del gasto de consumo de una familia pueden muy bien inducir a otra familia a aumentar su gasto de consumo para no quedar rezagada.

Antes de encontrar la razón de la autocorrelación es esencial aclarar la terminología. Aunque hoy en día es común tratar como sinónimos los términos **autocorrelación** y **correlación serial**, algunos autores prefieren diferenciarlos. Por ejemplo, Tintner define autocorrelación como "correlación rezagada de una serie dada consigo misma, rezagada por un número de unidades de tiempo", mientras que reserva el término correlación serial para "correlación rezagada entre

[1] Al respecto, véase William H. Greene, *Econometric Analysis*, 4a. ed., Prentice Hall, Nueva Jersey, 2000, cap. 11, y Paul A. Rudd, *An Introduction to Classical Econometric Theory*, Oxford University Press, 2000, cap. 19.

[2] Maurice G. Kendall y William R. Buckland, *A Dictionary of Statistical Terms*, Hafner Publishing Company, Nueva York, 1971, p. 8.

dos series diferentes”.[3] Así, la correlación entre dos series de tiempo como $u_1, u_2, \ldots, u_{10}$ y $u_2, u_3, \ldots, u_{11}$, donde la primera es igual a la última rezagada un periodo, es autocorrelación, mientras que la correlación entre dos series de tiempo como $u_1, u_2, \ldots, u_{10}$ y $v_2, v_3, \ldots, v_{11}$, donde $u$ y $v$ son dos series de tiempo diferentes, se denomina correlación serial. Aunque la distinción entre ambos puede ser útil, en este libro los consideraremos sinónimos.

Visualicemos algunos patrones razonables de autocorrelación y de no autocorrelación de la figura 12.1. Las figuras 12.1*a*) a *d*) muestran un patrón distinguible entre las $u$. La figura 12.1*a*) muestra un patrón cíclico; las figuras 12.1*b*) y *c*) sugieren una tendencia lineal hacia arriba o hacia abajo en las perturbaciones; y la figura 12.1*d*) indica que hay términos de tendencia tanto lineal como cuadrática en las perturbaciones. Sólo la figura 12.1*e*) indica que no hay un patrón sistemático, y apoya así el supuesto de no autocorrelación del modelo clásico de regresión lineal.

La pregunta natural es: ¿por qué ocurre la correlación serial? Hay diversas razones, algunas de las cuales son las siguientes:

### *Inercia*

Una característica relevante de la mayoría de las series de tiempo económicas es la inercia o pasividad. Como bien se sabe, las series de tiempo como PNB, índices de precios, producción, empleo y desempleo presentan ciclos (económicos). A partir del fondo de la recesión, cuando se inicia la recuperación económica, la mayoría de estas series empieza a moverse hacia arriba. En este movimiento ascendente, el valor de una serie en un punto del tiempo es mayor que su valor anterior. Así, se genera un “impulso” en ellas, y continuará hasta que suceda otra cosa (por ejemplo, un aumento en la tasa de interés o en los impuestos, o ambos) para reducirlo. Por consiguiente, es probable que, en las regresiones que consideran datos de series de tiempo, las observaciones sucesivas sean interdependientes.

### *Sesgo de especificación: caso de variables excluidas*

En el análisis empírico, con frecuencia el investigador empieza con un modelo de regresión razonable que puede no ser “perfecto”. Después del análisis de regresión, el investigador haría el examen *post mortem* para ver si los resultados coinciden con las expectativas *a priori*. De no ser así, iniciaría “la cirugía”. Por ejemplo, el investigador graficaría los residuos $\hat{u}_i$ obtenidos de la regresión ajustada y observaría patrones como los de las figuras 12.1*a*) a *d*). Estos residuos (representaciones de las $u_i$) pueden sugerir la inclusión de algunas variables originalmente candidatas pero que no se incluyeron en el modelo por diversas razones. Es el caso del sesgo de especificación ocasionado por **variables excluidas**. Con frecuencia, la inclusión de tales variables elimina el patrón de correlación observado entre los residuales. Por ejemplo, suponga que tenemos el siguiente modelo de demanda:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad \textbf{(12.1.2)}$$

donde $Y$ = cantidad de carne de res demandada, $X_2$ = precio de la carne de res, $X_3$ = ingreso del consumidor, $X_4$ = precio del cerdo y $t$ = tiempo.[4] Sin embargo, por alguna razón efectuamos la siguiente regresión:

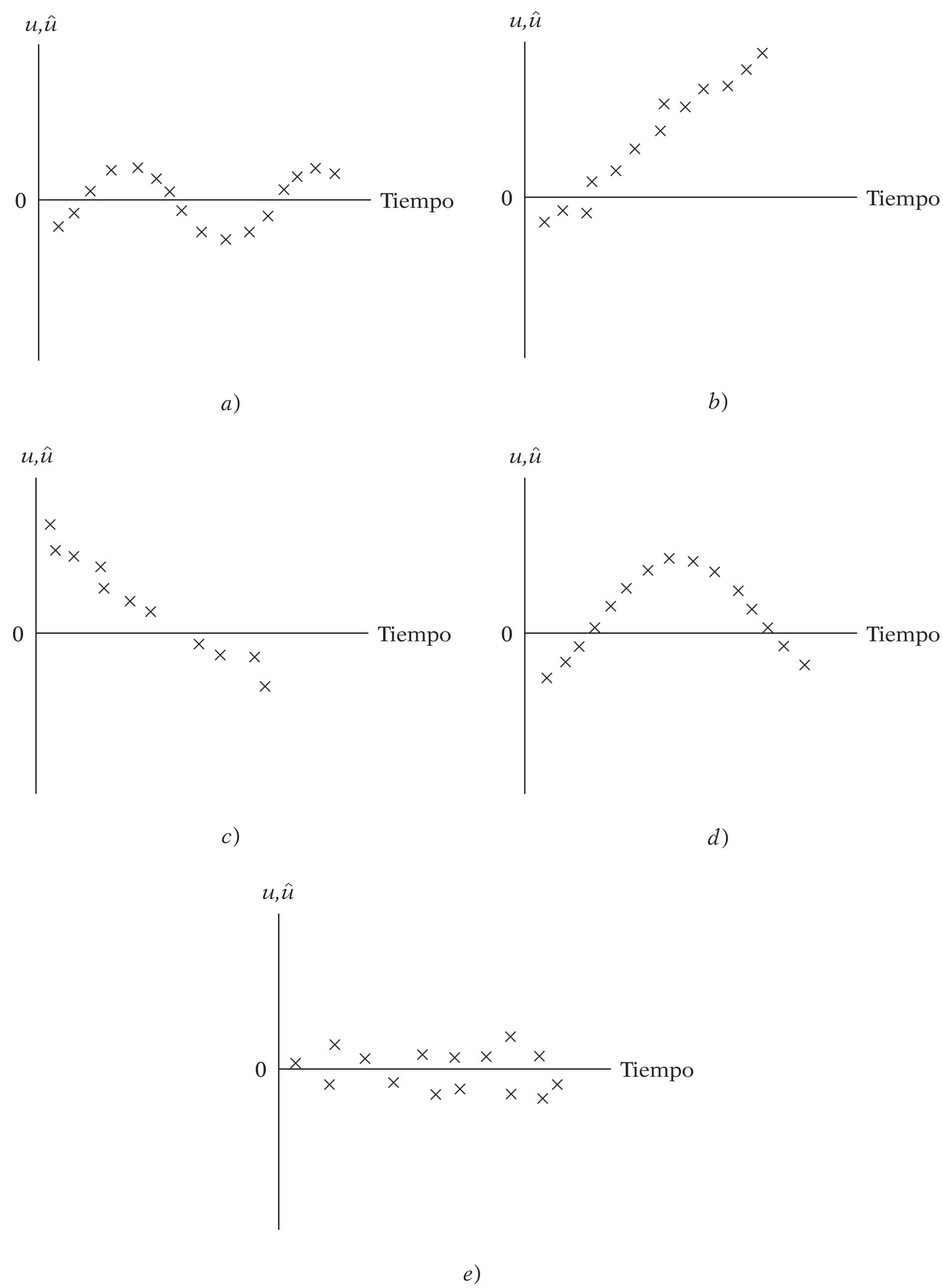
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad \textbf{(12.1.3)}$$

Ahora, si (12.1.2) es el modelo “correcto”, el “verdadero” o la relación verdadera, efectuar (12.1.3) equivale a permitir que $v_t = \beta_4 X_{4t} + u_t$. Así, en la medida en que el precio del cerdo afecte el consumo de carne de res, el término de error o de perturbación $v$ reflejará un patrón sistemático,

[3] Gerhard Tintner, *Econometrics*, John Wiley & Sons, Nueva York, 1965.

[4] Por convención, utilizaremos el subíndice $t$ para denotar información de series de tiempo y el subíndice usual $i$ para datos de corte transversal.

**FIGURA 12.1**
Patrones de autocorrelación y no autocorrelación.




lo que crea (una falsa) autocorrelación. Una prueba sencilla de esto sería llevar a cabo (12.1.2) y (12.1.3) y ver si la autocorrelación observada en el modelo (12.1.3), de existir, desaparece cuando se efectúa (12.1.2).[5] Analizaremos el mecanismo real para detectar la autocorrelación en la sección 12.6, donde mostraremos que una gráfica de los residuos de las regresiones (12.1.2) y (12.1.3) con frecuencia aclara en gran medida el problema de correlación serial.

[5] Si se encuentra que el problema real es de sesgo de especificación y no de autocorrelación, entonces, como veremos en el capítulo 13, los estimadores de MCO de los parámetros en la ecuación (12.1.3) pueden ser sesgados e inconsistentes.

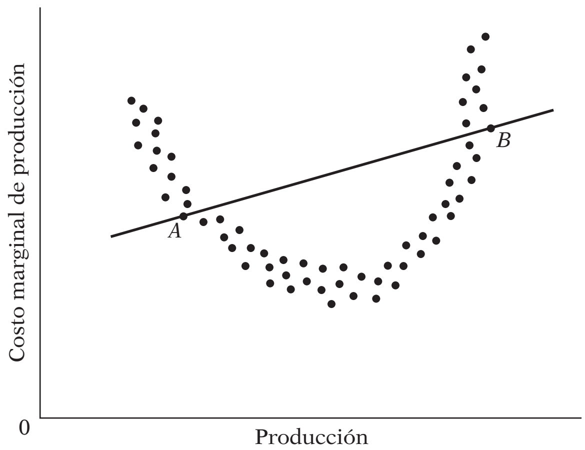



**FIGURA 12.2**
Sesgo de especificación: Forma funcional incorrecta.

### *Sesgo de especificación: forma funcional incorrecta*

Suponga que el modelo "verdadero" o correcto en un estudio de costo-producción es el siguiente:

$$\text{Costo marginal}_i = \beta_1 + \beta_2 \text{ producción}_i + \beta_3 \text{ producción}_i^2 + u_i \tag{12.1.4}$$

pero ajustamos el siguiente modelo:

$$\text{Costo marginal}_i = \alpha_1 + \alpha_2 \text{ producción}_i + v_i \tag{12.1.5}$$

La curva de costo marginal correspondiente al "verdadero" modelo se muestra en la figura 12.2, junto con la curva de costo lineal "incorrecta".

Como se muestra en la figura 12.2, entre los puntos $A$ y $B$ la curva de costo marginal lineal sobreestimará consistentemente el costo marginal verdadero, mientras que más allá de estos puntos, lo subestimará consistentemente. Este resultado es de esperarse porque el término de perturbación $v_i$ es, en realidad, igual a producción$^2 + u_i$, y, por tanto, capta el efecto sistemático del término producción$^2$ sobre el costo marginal. En este caso, $v_i$ reflejará autocorrelación por el uso de una forma funcional incorrecta. En el capítulo 13 consideraremos diversos métodos para detectar sesgos de especificación.

### *Fenómeno de la telaraña*

La oferta de muchos productos agrícolas refleja el llamado fenómeno de la telaraña, en donde la oferta reacciona al precio con un rezago de un periodo debido a que la instrumentación de las decisiones de oferta tarda algún tiempo (periodo de gestación). Por tanto, en la siembra de cultivos al principio de año, los agricultores reciben influencia del precio prevaleciente el año anterior, de forma que su función de oferta es

$$\text{Oferta}_t = \beta_1 + \beta_2 P_{t-1} + u_t \tag{12.1.6}$$

Suponga que al final del periodo $t$, el precio $P_t$ resulta inferior a $P_{t-1}$. Por consiguiente, es muy probable que en el periodo $t + 1$ los agricultores decidan producir menos de lo que produjeron en el periodo $t$. Obvio, en esta situación no esperaremos que las perturbaciones $u_i$ sean aleatorias, porque si los agricultores producen excedentes en el año $t$, es probable que reduzcan su producción en $t + 1$, y así sucesivamente, para generar un patrón de telaraña.

### *Rezagos*

En una regresión de series de tiempo del gasto de consumo sobre el ingreso no es extraño encontrar que el gasto de consumo en el periodo actual dependa, entre otras cosas, del gasto de consumo del periodo anterior.

Es decir,

$$\text{Consumo}_t = \beta_1 + \beta_2 \,\text{ingreso}_t + \beta_3 \,\text{consumo}_{t-1} + u_t \qquad (12.1.7)$$

Una regresión como (12.1.7) se conoce como **autorregresión** porque una variable explicativa es el valor rezagado de la variable dependiente. (Estudiaremos estos modelos en el capítulo 17.) El razonamiento de un modelo como (12.1.7) es sencillo. Los consumidores no cambian sus hábitos de consumo fácilmente por razones psicológicas, tecnológicas o institucionales. Ahora, si ignoramos el término rezagado en (12.1.7), el término de error resultante reflejará un patrón sistemático debido a la influencia del consumo rezagado en el consumo actual.

### *"Manipulación" de datos*

En el análisis empírico con frecuencia se "manipulan" los datos simples. Por ejemplo, en las regresiones de series de tiempo con datos trimestrales, por lo general estos datos provienen de datos mensuales a los que se agregan simplemente las observaciones de tres meses y se divide la suma entre 3. Este procedimiento de promediar las cifras suaviza en cierto grado los datos al eliminar las fluctuaciones en los datos mensuales. Por consiguiente, la gráfica referente a datos trimestrales aparece mucho más suave que la que contiene los datos mensuales, y este suavizamiento puede, por sí mismo, inducir un patrón sistemático en las perturbaciones, lo que agrega autocorrelación. Otra fuente de manipulación es la **interpolación** o **extrapolación** de datos. Por ejemplo, el Censo de Población se realiza cada 10 años en Estados Unidos, y los dos últimos se efectuaron en 1990 y 2000. Ahora bien, si necesitamos datos para algún año comprendido en el periodo intercensal, la práctica común consiste en interpolar con base en algunos supuestos *ad hoc*. Todas estas técnicas de "manejo" podrían imponer sobre los datos un patrón sistemático que quizá no estaría presente en los datos originales.[6]

### *Transformación de datos*

Como un ejemplo, considere el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (12.1.8)$$

donde, digamos, $Y$ = gasto de consumo y $X$ = ingreso. Como (12.1.8) es válida para cada periodo, también lo es para el periodo anterior $(t-1)$. Así, podemos expresar (12.1.8) como

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (12.1.9)$$

$Y_{t-1}$, $X_{t-1}$ y $u_{t-1}$ se conocen como los **valores rezagados** de $Y$, $X$ y $u$, respectivamente; en este caso están rezagados un periodo. Más tarde, en este mismo capítulo y en varias partes del libro, veremos la importancia de dichos valores rezagados.

Ahora bien, si restamos (12.1.9) de (12.1.8), obtenemos

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \qquad (12.1.10)$$

donde $\Delta$, llamado **operador de primeras diferencias**, indica que se toman diferencias sucesivas de las variables en cuestión. Por tanto, $\Delta Y_t = (Y_t - Y_{t-1})$, $\Delta X_t = (X_t - X_{t-1})$ y $\Delta u_t = (u_t - u_{t-1})$. Para propósitos empíricos, escribimos (12.1.10) como

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \qquad (12.1.11)$$

donde $v_t = \Delta u_t = (u_t - u_{t-1})$.

[6] Al respecto, véase William H. Greene, *op. cit.*, p. 526.

La ecuación (12.1.9) se conoce como la **forma de nivel,** y la ecuación (12.1.10), como la **forma en (primeras) diferencias**. Ambas formas son frecuentes en el análisis empírico. Por ejemplo, si en (12.1.9) $Y$ y $X$ representan los logaritmos del gasto de consumo y el ingreso, entonces en (12.1.10) $\Delta Y$ y $\Delta X$ representarán los cambios en los logaritmos del gasto de consumo y del ingreso. Pero, como sabemos, un cambio en el logaritmo de una variable —si se multiplica por 100— es un cambio relativo, o un cambio porcentual. De modo que, en vez de estudiar las relaciones entre variables en la forma de nivel, podemos interesarnos por las relaciones en la forma de crecimiento.

Ahora bien, si el término de error en (12.1.8) satisface los supuestos usuales de los MCO, sobre todo el de inexistencia de autocorrelación, podemos probar que el término de error $v_t$ en (12.1.11) está autocorrelacionado. (La prueba se encuentra en el apéndice 12A, sección 12A.1.) Los modelos como (12.1.11) se denominan **modelos dinámicos de regresión**; es decir, son modelos con regresadas rezagadas. Estudiaremos con detalle estos modelos en el capítulo 17.

Lo importante del ejemplo anterior es que a veces la autocorrelación puede inducirse como resultado de transformar el modelo original.

### *No estacionariedad*

Mencionamos en el capítulo 1 que, al trabajar con datos de series de tiempo, quizá habría que averiguar si una determinada serie de tiempo es estacionaria. Aunque el tema de no estacionariedad se analiza con mayor detalle en los capítulos de econometría de series de tiempo de la **parte 5** del libro, una serie de tiempo es estacionaria, de manera informal, si sus características (por ejemplo, media, varianza y covarianza) son invariantes respecto del tiempo; es decir, no cambian en relación con el tiempo. Si no es así, tenemos una serie de tiempo no estacionaria.

Como veremos en la **parte 5**, en un modelo de regresión como (12.1.8) es muy probable que $Y$ y $X$ sean no estacionarias, y por consiguiente, que el error $u$ también sea no estacionario.[7] En ese caso, el término de error mostrará autocorrelación.

Así, en resumen, hay varias razones por las que el término de error en un modelo de regresión pueda estar autocorrelacionado. En lo que resta del capítulo investigaremos con cierto detalle los problemas planteados por la autocorrelación y lo que se puede hacer al respecto.

Cabe notar también que la autocorrelación puede ser positiva [figura 12.3*a*)] o negativa, aunque la mayoría de las series de tiempo económicas por lo general muestra autocorrelación positiva, pues casi todas se desplazan hacia arriba o hacia abajo en extensos periodos y no exhiben un movimiento ascendente y descendente constante, como el de la figura 12.3*b*).

## 12.2 Estimación de MCO en presencia de autocorrelación

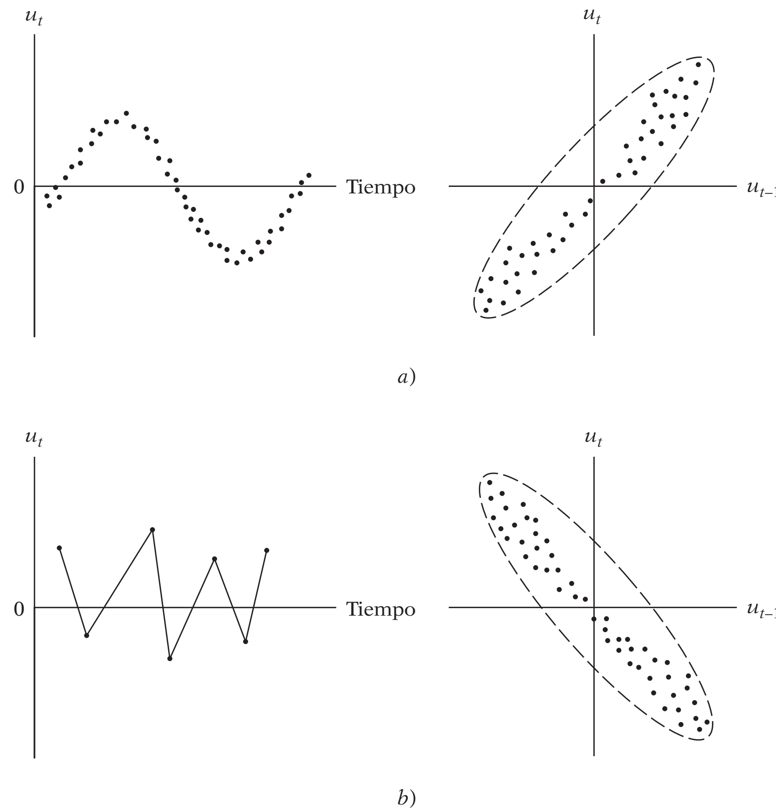
¿Qué sucede con los estimadores de MCO y sus varianzas si introducimos autocorrelación en las perturbaciones con la suposición de que $E(u_t u_{t+s}) \neq 0$ $(s \neq 0)$, pero conservamos todos los demás supuestos del modelo clásico?[8] Observe de nuevo que ahora utilizamos el subíndice $t$ en las perturbaciones para destacar que se trata de datos de series de tiempo.

Regresamos al modelo de regresión de dos variables para explicar sus ideas básicas, a saber, $Y_t = \beta_1 + \beta_2 X_t + u_t$. Para orientar el camino, ahora debemos suponer el mecanismo que generan las $u_t$, pues $E(u_t u_{t+s}) \neq 0$ $(s \neq 0)$ es muy general como supuesto para ser de alguna utilidad

[7] Como también veremos en la **parte 5**, aunque $X$ y $Y$ sean no estacionarias, es posible encontrar que $u$ lo sea. Más adelante analizaremos lo que esto implica.

[8] Si $s = 0$, obtenemos $E(u_t^2)$. Como $E(u_t) = 0$ por hipótesis, $E(u_t^2)$ representará la varianza del término de error, que obviamente es diferente de cero (¿por qué?).

**FIGURA 12.3**
*a*) Autocorrelación positiva y *b*) autocorrelación negativa.




práctica. Como punto de partida, o primera aproximación, podemos suponer que los términos de error o perturbación se generan de la siguiente manera:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \tag{12.2.1}$$

donde $\rho$ (= rho) se conoce como **coeficiente de autocovarianza** y $e_t$ es la perturbación estocástica establecida de forma que satisface los supuestos habituales de MCO, a saber,

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ \text{var}\,(\varepsilon_t) &= \sigma_\varepsilon^2 \\ \text{cov}\,(\varepsilon_t, \varepsilon_{t+s}) &= 0 \qquad s \neq 0 \end{aligned} \tag{12.2.2}$$

En los textos de ingeniería, un término de error con las propiedades anteriores a menudo se conoce como **término de error de ruido blanco**. Lo que (12.2.1) postula es que el valor del término de perturbación en el periodo $t$ es igual a $\rho$ multiplicada por su valor en el periodo anterior más un término de error puramente aleatorio.

El esquema (12.2.1) se conoce como **esquema autorregresivo de primer orden de Markov,** o simplemente **esquema autorregresivo de primer orden,** y suele denotarse como **AR(1)**. El nombre *autorregresivo* es apropiado porque (12.2.1) puede interpretarse como la regresión de $u_t$ sobre sí misma con un rezago de un periodo. Es de primer orden porque sólo participan $u_t$ y su valor pasado inmediato; es decir, el rezago máximo es 1. Si el modelo fuera $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$, sería un **AR(2)**, o esquema autorregresivo de segundo orden, y así sucesivamente. Estudiaremos esos esquemas de orden superior en la **parte 5**, en los capítulos sobre la econometría de series de tiempo.

A propósito, observe que $\rho$, el coeficiente de autocovarianza en (12.2.1), también se interpreta como el **coeficiente de autocorrelación de primer orden**, o, en forma más precisa, **coeficiente de autocorrelación del rezago 1**.[9]

Con el esquema AR(1) se demuestra que (véase el apéndice 12A, sección 12A.2):

$$\operatorname{var}(u_t) = E\left(u_t^2\right) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.3}$$

$$\operatorname{cov}(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.4}$$

$$\operatorname{cor}(u_t, u_{t+s}) = \rho^s \tag{12.2.5}$$

donde cov($u_t$, $u_{t+s}$) significa la covarianza entre los términos de error de $s$ periodos distantes, y cor($u_t$, $u_{t+s}$), la correlación entre los términos de error de $s$ periodos distantes. Note que, debido a la propiedad de simetría de las covarianzas y las correlaciones, cov($u_t$, $u_{t+s}$) = cov($u_t$, $u_{t-s}$) y cor($u_t$, $u_{t+s}$) = cor($u_t$, $u_{t-s}$).

Como $\rho$ es una constante con valor entre $-1$ y $+1$ (12.2.3), muestra que, con el esquema AR(1), la varianza de $u_t$ *es todavía homoscedástica*; no obstante, $u_t$ está correlacionada no sólo con su valor inmediato anterior, sino con sus valores de varios periodos anteriores. Resulta *crítico* notar que $|\rho| < 1$; es decir, el valor absoluto de $\rho$ es menor que 1. Si, por ejemplo, $\rho$ es igual a 1, las varianzas y covarianzas listadas antes no están definidas. Si $|\rho| < 1$, se dice que el proceso AR(1) en (12.2.1) es *estacionario*; es decir, la media, la varianza y la covarianza de $u_t$ no cambian respecto del tiempo. Si $|\rho|$ es menor que 1, resulta claro de (12.2.4) que el valor de la covarianza declinará conforme se retroceda al pasado distante. En breve veremos la utilidad de los resultados anteriores.

Una razón para utilizar el proceso AR(1) no es sólo su simplicidad en comparación con los esquemas AR de orden superior, sino porque es muy útil para muchas aplicaciones. Además, hay gran cantidad de trabajo teórico y empírico sobre el esquema AR(1).

Ahora regresemos al modelo de regresión con dos variables: $Y_t = \beta_1 + \beta_2 X_t + u_t$. Sabemos del capítulo 3 que el estimador de MCO del coeficiente de pendiente es

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{12.2.6}$$

y su varianza está dada por

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{12.2.7}$$

en donde las letras minúsculas denotan como siempre la desviación de sus valores medios.

[9] Este nombre se justifica fácilmente. Por definición, el coeficiente de correlación (poblacional) entre $u_t$ y $u_{t-1}$ es

$$\rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{\operatorname{var}(u_t)}\sqrt{\operatorname{var}(u_{t-1})}}$$

$$= \frac{E(u_t u_{t-1})}{\operatorname{var}(u_{t-1})}$$

porque $E(u_t) = 0$ para cada $t$ y var($u_t$) = var($u_{t-1}$), pues aún es válido el supuesto de homoscedasticidad. El lector puede ver que $\rho$ es también el coeficiente de pendiente en la regresión de $u_t$ sobre $u_{t-1}$.

Ahora, con el esquema AR(1), se muestra que la varianza de este estimador es:

$$\text{var}\,(\hat{\beta}_2)_{\text{AR1}} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right] \tag{12.2.8}$$

donde var$(\hat{\beta}_2)_{\text{AR1}}$ significa la varianza de $\hat{\beta}_2$ con el esquema autorregresivo de primer orden.

Una comparación de (12.2.8) con (12.2.7) muestra que la primera es igual a la última más un término que depende de $\rho$, así como de las autocorrelaciones muestrales entre los valores tomados por la regresora $X$ en varios rezagos.[10] Y no podemos predecir que la var$(\hat{\beta}_2)$ será menor o mayor que var$(\hat{\beta}_2)_{\text{AR1}}$ [véase ecuación (12.4.1), más adelante]. Por supuesto, si $\rho$ es cero, las dos fórmulas coincidirán, como debe ser (¿por qué?). Asimismo, si las correlaciones entre los valores sucesivos de la regresora son muy pequeñas, la varianza usual de MCO del estimador de pendiente no estará muy sesgada. Sin embargo, como principio general, las dos varianzas no serán iguales.

Para que se dé una idea respecto de la diferencia entre las varianzas dadas en (12.2.7) y (12.2.8), suponga que la regresora $X$ también sigue el esquema autorregresivo de primer orden con un coeficiente de autocorrelación de $r$. Se demuestra entonces que (12.2.8) se reduce a:

$$\text{var}\,(\hat{\beta}_2)_{\text{AR(1)}} = \frac{\sigma^2}{\sum x_t^2}\left(\frac{1 + r\rho}{1 - r\rho}\right) = \text{var}\,(\hat{\beta}_2)_{\text{MCO}}\left(\frac{1 + r\rho}{1 - r\rho}\right) \tag{12.2.9}$$

Si, por ejemplo, $r = 0.6$ y $\rho = 0.8$, con (12.2.9) podemos verificar que var$(\hat{\beta}_2)_{\text{AR1}} = 2.8461$ var$(\hat{\beta}_2)_{\text{MCO}}$. Para expresarlo de otra manera, var$(\hat{\beta}_2)_{\text{MCO}} = \frac{1}{2.8461}$ var$(\hat{\beta}_2)_{\text{AR1}} = 0.3513$ var$(\hat{\beta}_2)_{\text{AR1}}$. Es decir, la fórmula usual MCO [es decir, (12.2.7)] subestimará la varianza de $(\hat{\beta}_2)_{\text{AR1}}$ aproximadamente 65%. Como resulta obvio, esta respuesta es específica para los valores dados de $r$ y $\rho$. Pero lo importante de este ejercicio es advertir que aplicar a ciegas las fórmulas usuales de MCO para calcular las varianzas y los errores estándar de los estimadores de MCO pueden generar resultados muy erróneos.

Suponga que seguimos utilizando el estimador de MCO $\hat{\beta}_2$ y ajustamos la fórmula de varianza habitual, con el esquema AR(1). Es decir empleamos el $\hat{\beta}_2$ dado por (12.2.6) pero con la fórmula de varianza dada por (12.2.8). ¿Cúales son ahora las propiedades de $\hat{\beta}_2$? Es fácil probar que $\hat{\beta}_2$ es aún lineal e insesgado. En realidad, como se observa en el apéndice 3A, sección 3A.2, no se requiere el supuesto de no correlación serial ni el de no heteroscedasticidad para demostrar que $\hat{\beta}_2$ es insesgado. ¿Es $\hat{\beta}_2$ aún MELI? Por desgracia, no; en la clase de estimadores lineales e insesgados, no tiene varianza mínima. En resumen, aunque $\hat{\beta}_2$ es lineal e insesgado, no es eficiente (en términos relativos, por supuesto). El lector notará que este hallazgo es muy similar al de que $\hat{\beta}_2$ es menos eficiente en presencia de heteroscedasticidad. Allí vimos que el estimador eficiente era el estimador de mínimos cuadrados ponderados, $\hat{\beta}_2^*$, dado en (11.3.8), un caso especial del estimador de mínimos cuadrados generalizados (MCG). En el caso de autocorrelación, ¿podemos encontrar un estimador que sea MELI? La respuesta es sí, como veremos en la siguiente sección.

[10] Observe que el término $r = \sum x_t x_{t+1} / \sum x_t^2$ es la correlación entre $X_t$ y $X_{t+1}$ (o $X_{t-1}$, pues el coeficiente de correlación es simétrico); $r^2 = \sum x_t x_{t+2} / \sum x_t^2$ es la correlación entre las $X$ rezagadas dos periodos, etcétera.

## 12.3 Estimador MELI en presencia de autocorrelación

Al continuar con el modelo de dos variables y suponer el proceso AR(1), es posible demostrar que el estimador MELI de $\beta_2$ está dado por la siguiente expresión:[11]

$$\hat{\beta}_2^{\text{MCG}} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + C \tag{12.3.1}$$

donde $C$ es un factor de corrección que puede ignorarse en la práctica. Observe que el subíndice $t$ varía ahora de $t = 2$ a $t = n$. Y su varianza está dada por

$$\text{var}\,\hat{\beta}_2^{\text{MCG}} = \frac{\sigma^2}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + D \tag{12.3.2}$$

donde $D$ también es un factor de corrección que puede ignorarse en la práctica. (Véase el ejercicio 12.18.)

El estimador $\hat{\beta}_2^{\text{MCG}}$, como lo indica el superíndice, se obtiene por el método de MCG. Como mencionamos en el capítulo 11, en MCG se incorpora directamente cualquier información adicional que se tenga (por ejemplo, la naturaleza de la heteroscedasticidad o de la autocorrelación) en el proceso de estimación mediante la transformación de variables, mientras que en MCO tal información adicional no se considera directamente. Como puede ver, el estimador de MCG de $\beta_2$ dado en (12.3.1) incorpora el parámetro de autocorrelación $\rho$ en la fórmula de estimación, mientras que la fórmula de MCO dada en (12.2.6) simplemente lo ignora. La intuición indica que por esta razón el estimador de MCG es MELI y el estimador de MCO no lo es; el estimador de MCG emplea al máximo la información disponible.[12] No es preciso mencionar que si $\rho = 0$, no hay información adicional que deba considerarse y, por tanto, los estimadores de MCG y MCO son idénticos.

*En resumen*, con autocorrelación, el estimador de MCG dado en (12.3.1) es MELI y la varianza mínima está dada ahora por (12.3.2) y no por (12.2.6), ni, obvio, por (12.2.7).

### *Nota técnica*

Como observamos en el capítulo anterior, el teorema de Gauss-Markov proporciona sólo la condición suficiente para que los MCO sean MELI. Las condiciones suficientes y necesarias para que los MCO sean MELI las establece el **teorema de Kruskal**, mencionado en el capítulo anterior. Así, en algunos casos los MCO pueden ser MELI, a pesar de la autocorrelación. Pero tales casos son poco frecuentes en la práctica.

¿Qué sucede si se continúa trabajando despreocupadamente con el procedimiento MCO usual, a pesar de la autocorrelación? La respuesta se da en la siguiente sección.

---

[11] Para las demostraciones, véase Jan Kmenta, *Elements of Econometrics*, Macmillan, Nueva York, 1971, pp. 274-275. El factor de corrección $C$ se refiere a la primera observación ($Y_1$, $X_1$). Sobre este punto, véase el ejercicio 12.18.

[12] La prueba formal de que $\hat{\beta}_2^{\text{MCG}}$ es MELI se encuentra en Kmenta, *ibid.* Sin embargo, la tediosa prueba algebraica se simplifica considerablemente mediante notación matricial. Véase J. Johnston, *Econometric Methods*, 3a. ed., McGraw-Hill, Nueva York, 1984, pp. 291-293.

## 12.4 Consecuencias de utilizar MCO en presencia de autocorrelación

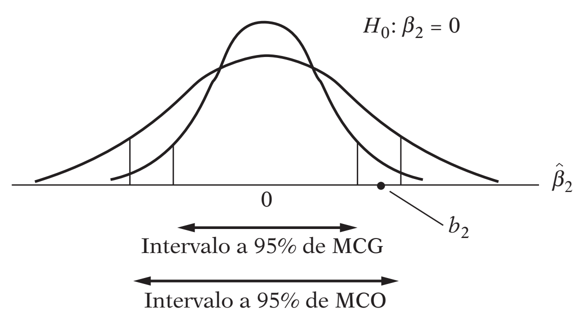
Como en la heteroscedasticidad, en presencia de autocorrelación los estimadores continúan siendo lineales e insesgados, al igual que consistentes, y están distribuidos de forma asintóticamente normal, pero dejan de ser eficientes (es decir, no tienen varianza mínima). ¿Qué sucede entonces con los procedimientos usuales de pruebas de hipótesis si se conservan los estimadores de MCO? De nuevo, como en el caso de heteroscedasticidad, se distinguen dos casos. Por razones pedagógicas continuaremos trabajando con el modelo de dos variables, aunque el siguiente análisis puede extenderse a regresiones múltiples sin mucho esfuerzo.[13]

### Estimación por MCO tomando en cuenta la autocorrelación

Como se mencionó, $\hat{\beta}_2$ no es MELI, y aunque se fuera a usar $\text{var}(\hat{\beta}_2)_{\text{AR1}}$, es probable que los intervalos de confianza derivados de allí sean más amplios que los basados en el procedimiento MCG. Como señala Kmenta, es probable que éste sea el resultado aunque el tamaño de la muestra se incremente indefinidamente.[14] Es decir, $\hat{\beta}_2$ *no es asintóticamente eficiente*. La implicación de este hallazgo para pruebas de hipótesis es clara: es probable que se declare un coeficiente estadísticamente no significativo (es decir, no diferente de cero) aunque en realidad pueda serlo (es decir, si se basa en el procedimiento MCG correcto). Esta diferencia se ve claramente en la figura 12.4. En ella se muestran intervalos de confianza a 95% de MCO [AR(1)] y MCG suponiendo que el verdadero $\beta_2 = 0$. Considere una estimación particular de $\beta_2$, por ejemplo, $b_2$. Como $b_2$ cae en el intervalo de confianza de MCO, podemos aceptar la hipótesis de que el verdadero $\beta_2$ es cero con 95% de confianza. Pero si utilizáramos el intervalo de confianza de MCG (correcto), podríamos rechazar la hipótesis nula de que el verdadero $\beta_2$ es cero, pues $b_2$ cae en la región de rechazo.

**El mensaje es: para establecer intervalos de confianza y probar hipótesis, debe utilizarse MCG y no MCO, aunque los estimadores derivados de este último sean insesgados y consistentes.** (No obstante, véase la sección 12.11.)

**FIGURA 12.4**
Intervalos de confianza a 95% de MCG y MCO.




### Estimación por MCO ignorando la autocorrelación

La situación es potencialmente muy grave si no sólo utilizamos $\hat{\beta}_2$ sino también $\text{var}(\hat{\beta}_2) = \sigma^2/\sum x_t^2$, con lo cual se ignora por completo el problema de autocorrelación; es decir, creemos erróneamente que los supuestos usuales del modelo clásico se mantienen. Surgirán errores por las siguientes razones:

1. Es probable que la varianza de los residuos $\hat{\sigma}^2 = \sum \hat{u}_t^2/(n-2)$ subestime la verdadera $\sigma^2$.
2. Como resultado, es probable que se sobreestime $R^2$.

[13] Pero el álgebra matricial se convierte casi en una necesidad para evitar tediosas manipulaciones algebraicas.

[14] Véase Kmenta, *op. cit.*, pp. 277-278.

3. Aunque $\sigma^2$ no esté subestimada, var($\hat{\beta}_2$) puede subestimar var($\hat{\beta}_2$)$_{\text{AR1}}$ [ecuación (12.2.8)], su varianza con autocorrelación (de primer orden), pese a que esta última sea ineficiente comparada con var($\hat{\beta}_2$)$^{\text{MCG}}$.
4. Por consiguiente, las pruebas de significancia $t$ y $F$ usuales dejan de ser válidas y, de aplicarse, es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los coeficientes de regresión estimados.

Para establecer algunas de estas proposiciones, regresemos al modelo de dos variables. Sabemos, del capítulo 3, que según el supuesto clásico

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)}$$

constituye un estimador insesgado de $\sigma^2$, es decir, $E(\hat{\sigma}^2) = \sigma^2$. Pero si hay autocorrelación, dada por AR(1), se ve que

$$E(\hat{\sigma}^2) = \frac{\sigma^2\{n - [2/(1-\rho)] - 2\rho r\}}{n-2} \tag{12.4.1}$$

donde $r = \sum_{t=1}^{n-1} x_t x_{t-1} / \sum_{t=1}^{n} x_t^2$, que puede interpretarse como el coeficiente de correlación (muestral) entre valores sucesivos de las $X$.[15] Si $\rho$ y $r$ son positivos (lo cual no es un supuesto improbable para la mayoría de las series de tiempo económicas), es claro de (12.4.1) que $E(\hat{\sigma}^2) < \sigma^2$; es decir, la fórmula usual de varianza residual, en promedio, subestimará la verdadera $\sigma^2$. En otras palabras, $\hat{\sigma}^2$ estará sesgada hacia abajo. Sobra decir que este sesgo en $\hat{\sigma}^2$ se transmitirá a var($\hat{\beta}_2$), pues, en la práctica, esta última se estima con la fórmula $\hat{\sigma}^2 / \sum x_t^2$.

Sin embargo, aunque $\sigma^2$ no se subestime, var($\hat{\beta}_2$) es un estimador *sesgado* de var($\hat{\beta}_2$)$_{\text{AR1}}$, lo cual se ve fácilmente al comparar (12.2.7) con (12.2.8),[16] pues las dos fórmulas no son iguales. En realidad, si $\rho$ es positivo (lo cual sucede en la mayoría de las series de tiempo económicas) y las $X$ están correlacionadas positivamente (que también sucede en la mayoría de las series de tiempo económicas), es claro que

$$\text{var}(\hat{\beta}_2) < \text{var}(\hat{\beta}_2)_{\text{AR1}} \tag{12.4.2}$$

es decir, la varianza de MCO usual de $\hat{\beta}_2$ subestima su varianza con AR(1) [véase la ecuación (12.2.9)]. Por consiguiente, si utilizamos var($\hat{\beta}_2$), inflaremos la precisión o exactitud (es decir, subestimamos el error estándar) del estimador $\hat{\beta}_2$. Como resultado, al calcular la razón $t$ como $t = \hat{\beta}_2/\text{ee}(\hat{\beta}_2)$ (según la hipótesis de que $\beta_2 = 0$), sobreestimamos el valor de $t$, y, por tanto, la significancia estadística de $\beta_2$ estimado. La situación tiende a empeorar si además $\sigma^2$ está subestimada, como ya observamos.

Para ver la forma en que quizá MCO subestime $\sigma^2$ y la varianza de $\hat{\beta}_2$, realicemos el siguiente **experimento Monte Carlo**. Suponga que en el modelo de dos variables "conocemos" las verdaderas $\beta_1 = 1$ y $\beta_2 = 0.8$. Por consiguiente, la FRP estocástica es

$$Y_t = 1.0 + 0.8X_t + u_t \tag{12.4.3}$$

[15] Véase S.M. Goldfeld y R.E. Quandt, *Nonlinear Methods in Econometrics*, North Holland Publishing Company, Ámsterdam, 1972, p. 183. A propósito, observe que si los errores están autocorrelacionados positivamente, el valor $R^2$ tiende a tener un sesgo hacia arriba, es decir, tiende a ser más grande que el $R^2$ en ausencia de tal correlación.

[16] Para una prueba formal, véase Kmenta, *op. cit.*, p. 281.

**TABLA 12.1**
**Ejemplo hipotético de términos de error positivamente autocorrelacionados**

| | $\varepsilon_t$ | $u_t = 0.7u_{t-1} + \varepsilon_t$ |
|---|---|---|
| 0 | 0 | $u_0 = 5$ (supuesto) |
| 1 | 0.464 | $u_1 = 0.7(5) + 0.464 = 3.964$ |
| 2 | 2.026 | $u_2 = 0.7(3.964) + 2.0262 = 4.8008$ |
| 3 | 2.455 | $u_3 = 0.7(4.8010) + 2.455 = 5.8157$ |
| 4 | −0.323 | $u_4 = 0.7(5.8157) - 0.323 = 3.7480$ |
| 5 | −0.068 | $u_5 = 0.7(3.7480) - 0.068 = 2.5556$ |
| 6 | 0.296 | $u_6 = 0.7(2.5556) + 0.296 = 2.0849$ |
| 7 | −0.288 | $u_7 = 0.7(2.0849) - 0.288 = 1.1714$ |
| 8 | 1.298 | $u_8 = 0.7(1.1714) + 1.298 = 2.1180$ |
| 9 | 0.241 | $u_9 = 0.7(2.1180) + 0.241 = 1.7236$ |
| 10 | −0.957 | $u_{10} = 0.7(1.7236) - 0.957 = 0.2495$ |

*Nota:* Los datos de $\varepsilon_t$ provienen de *A Million Random Digits and One Hundred Thousand Deviates*, Rand Corporation, Santa Mónica, California, 1950.

Por tanto,

$$E(Y_t \mid X_t) = 1.0 + 0.8X_t \tag{12.4.4}$$

lo cual da la verdadera línea de regresión poblacional. Supongamos que las $u_t$ se generan mediante el esquema autorregresivo de primer orden de la siguiente manera:

$$u_t = 0.7u_{t-1} + \varepsilon_t \tag{12.4.5}$$

donde $\varepsilon_t$ satisface todos los supuestos de MCO. Además, por conveniencia, supongamos que los $\varepsilon_t$ están normalmente distribuidos con media cero y varianza unitaria (= 1). La ecuación (12.4.5) postula que las perturbaciones consecutivas están correlacionadas positivamente, con un coeficiente de autocorrelación de +0.7, un grado más bien alto de dependencia.

Ahora, con una tabla de números aleatorios normales con media cero y varianza unitaria, generamos los 10 números aleatorios de la tabla 12.1, mediante el esquema (12.4.5) generamos $u_t$. Para iniciar el esquema necesitamos especificar el valor inicial de $u$, por ejemplo, $u_0 = 5$.

Al graficar las $u_t$ generadas en la tabla 12.1 obtenemos la figura 12.5, la cual muestra que, al principio, cada $u_t$ sucesiva tiene un valor más alto que su valor anterior, y después, por lo regular, es menor que su valor anterior, para mostrar, en general, una autocorrelación positiva.

Ahora suponga que los valores de $X$ están dados en 1, 2, 3, . . . , 10. Entonces, con estas $X$, podemos generar una muestra de 10 valores $Y$ de (12.4.3) y de los valores de $u_t$ dados en la tabla 12.1. Los detalles se presentan en la tabla 12.2. Si con los datos de la tabla 12.2, efectuamos la regresión de $Y$ sobre $X$, obtenemos la siguiente regresión (muestral):

$$\begin{aligned} \hat{Y}_t = {} & 6.5452 + 0.3051X_t \\ & (0.6153) \quad (0.0992) \\ t = {} & (10.6366) \quad (3.0763) \\ & r^2 = 0.5419 \qquad \hat{\sigma}^2 = 0.8114 \end{aligned} \tag{12.4.6}$$

mientras que la verdadera línea de regresión es como la dada por (12.4.4). Ambas líneas de regresión se muestran en la figura 12.6, la cual señala claramente cuánto distorsiona la línea de regresión ajustada a la verdadera línea de regresión; ésta subestima en gran medida al verdadero coeficiente de pendiente pero sobreestima al intercepto verdadero. (Sin embargo, observe que los estimadores de MCO aún son insesgados.)

La figura 12.6 también muestra por qué es probable que la verdadera varianza de $u_i$ esté subestimada por el estimador $\hat{\sigma}^2$, el cual se calcula a partir de las $\hat{u}_i$. Las $\hat{u}_i$ suelen estar cerca de

**FIGURA 12.5**
Correlación generada por el esquema $u_t = 0.7u_{t-1} + \varepsilon_t$ (tabla 12.1).

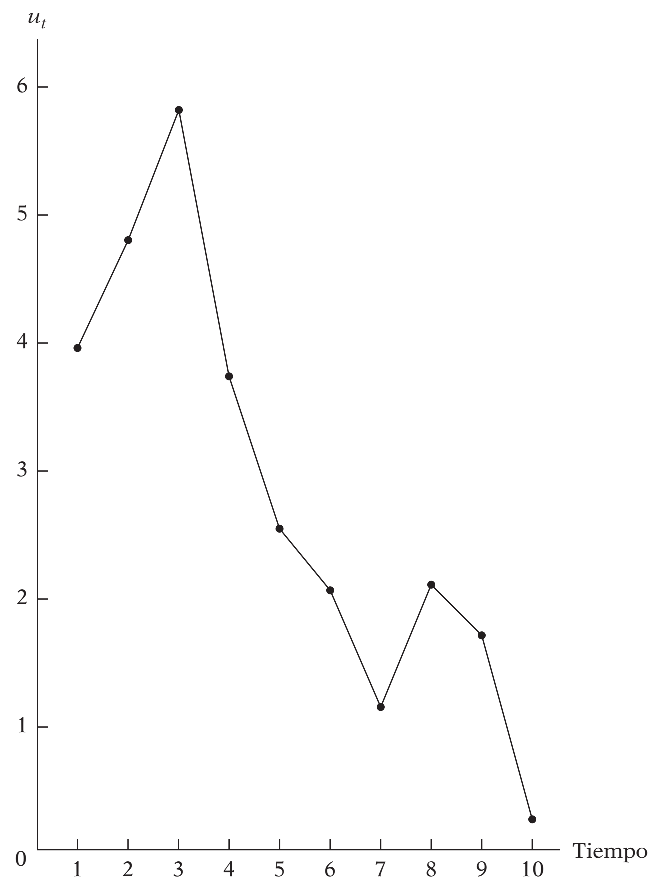



**TABLA 12.2**
**Generación de valores muestrales de *Y***

| $X_t$ | $u_t$ | $Y_t = 1.0 + 0.8X_t + u_t$ |
|---|---|---|
| 1 | 3.9640 | $Y_1 = 1.0 + 0.8(1) + 3.9640 = 5.7640$ |
| 2 | 4.8010 | $Y_2 = 1.0 + 0.8(2) + 4.8008 = 7.4008$ |
| 3 | 5.8157 | $Y_3 = 1.0 + 0.8(3) + 5.8157 = 9.2157$ |
| 4 | 3.7480 | $Y_4 = 1.0 + 0.8(4) + 3.7480 = 7.9480$ |
| 5 | 2.5556 | $Y_5 = 1.0 + 0.8(5) + 2.5556 = 7.5556$ |
| 6 | 2.0849 | $Y_6 = 1.0 + 0.8(6) + 2.0849 = 7.8849$ |
| 7 | 1.1714 | $Y_7 = 1.0 + 0.8(7) + 1.1714 = 7.7714$ |
| 8 | 2.1180 | $Y_8 = 1.0 + 0.8(8) + 2.1180 = 9.5180$ |
| 9 | 1.7236 | $Y_9 = 1.0 + 0.8(9) + 1.7236 = 9.9236$ |
| 10 | 0.2495 | $Y_{10} = 1.0 + 0.8(10) + 0.2495 = 9.2495$ |

*Nota:* Los datos de $u_t$ provienen de la tabla 12.1.

la línea ajustada (lo cual se debe al procedimiento MCO) pero se desvían sustancialmente de la verdadera FRP. Por tanto, no dan una imagen correcta de $u_i$. Para tener una mejor idea del grado en que se subestima el verdadero valor de $\sigma^2$ suponga que efectuamos otro experimento de muestreo. Con los valores de $X_t$ y $\varepsilon_t$ de las tablas 12.1 y 12.2, suponga que $\rho = 0$, es decir, que no hay autocorrelación. La nueva muestra de valores de $Y$, así generados, está en la tabla 12.3.

**FIGURA 12.6**
FRP verdadera y línea de regresión estimada para los datos de la tabla 12.2.

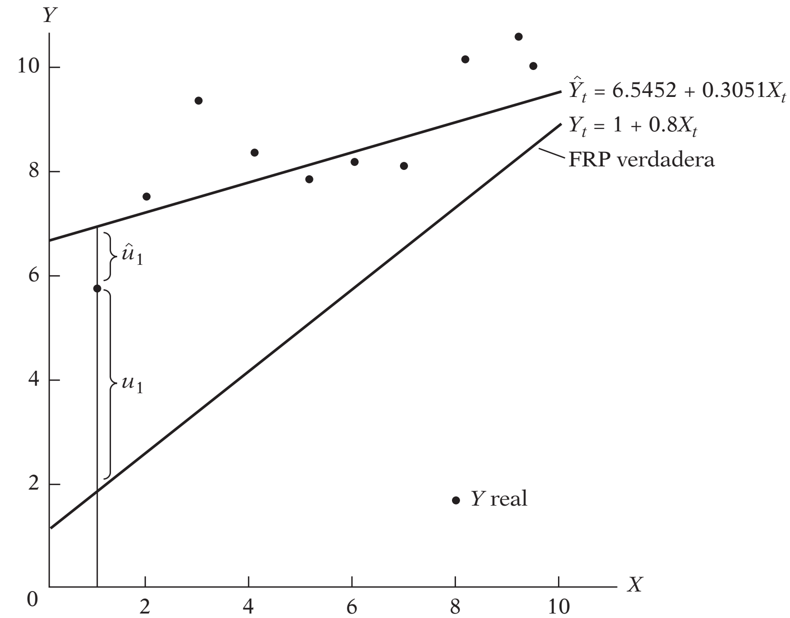



**TABLA 12.3**
**Muestra de valores *Y* con correlación serial igual a cero**

| $X_t$ | $\varepsilon_t = u_t$ | $Y_t = 1.0 + 0.8X_t + \varepsilon_t$ |
|---|---|---|
| 1 | 0.464 | 2.264 |
| 2 | 2.026 | 4.626 |
| 3 | 2.455 | 5.855 |
| 4 | −0.323 | 3.877 |
| 5 | −0.068 | 4.932 |
| 6 | 0.296 | 6.096 |
| 7 | −0.288 | 6.312 |
| 8 | 1.298 | 8.698 |
| 9 | 0.241 | 8.441 |
| 10 | −0.957 | 8.043 |

*Nota:* Como no hay autocorrelación, las $u_t$ y $\varepsilon_t$ son idénticas. Las $\varepsilon_t$ se tomaron de la tabla 12.1.

La regresión basada en la tabla 12.3 es la siguiente:

$$\begin{aligned} \hat{Y}_t &= 2.5345 + 0.6145X_t \\ &\quad\ (0.6796) \quad (0.1087) \\ t &= (3.7910) \quad (5.6541) \\ &\qquad\qquad r^2 = 0.7997 \qquad \hat{\sigma}^2 = 0.9752 \end{aligned} \tag{12.4.7}$$

Esta regresión es mucho más cercana a la "verdadera" porque las *Y* son ahora en esencia aleatorias. Observe que $\hat{\sigma}^2$ aumentó de 0.8114 ($\rho = 0.7$) a 0.975 ($\rho = 0$), y también que aumentaron los errores estándar de $\hat{\beta}_1$ y $\hat{\beta}_2$. Este resultado coincide con los resultados teóricos ya considerados.

## 12.5 Relación entre salarios y productividad en el sector de negocios de Estados Unidos, 1960-2005

Ahora que analizamos las consecuencias de la autocorrelación, la pregunta obvia es: ¿cómo detectarla y corregirla? Antes de adentrarnos en esos temas, consideremos un ejemplo concreto. La tabla 12.4 proporciona información sobre índices de remuneración real por hora $Y$ (RCOMPB) y producción por hora $X$ (PRODB) en el sector de negocios de la economía de Estados Unidos de 1960 a 2005; la base de los índices es 1992 = 100.

Primero graficamos los datos de $Y$ y $X$, con lo que obtenemos la figura 12.7. Como se espera que la relación entre la remuneración real y la productividad del trabajo sea positiva, no sorprende que las dos variables estén positivamente relacionadas. Lo que sí sorprende es que su relación sea casi lineal, aunque se da alguna pista de que, en niveles más altos de productividad, la relación entre ambas podría ser menos lineal. En consecuencia, decidimos estimar un modelo lineal y un log-lineal, con los siguientes resultados:

$$\begin{aligned} \hat{Y}_t &= \quad 32.7419 + \quad 0.6704X_t \\ \text{ee} &= \quad (1.3940) \qquad (0.0157) \\ t &= (23.4874) \quad (42.7813) \end{aligned} \tag{12.5.1}$$

$$r^2 = 0.9765 \qquad d = 0.1739 \qquad \hat{\sigma} = 2.3845$$

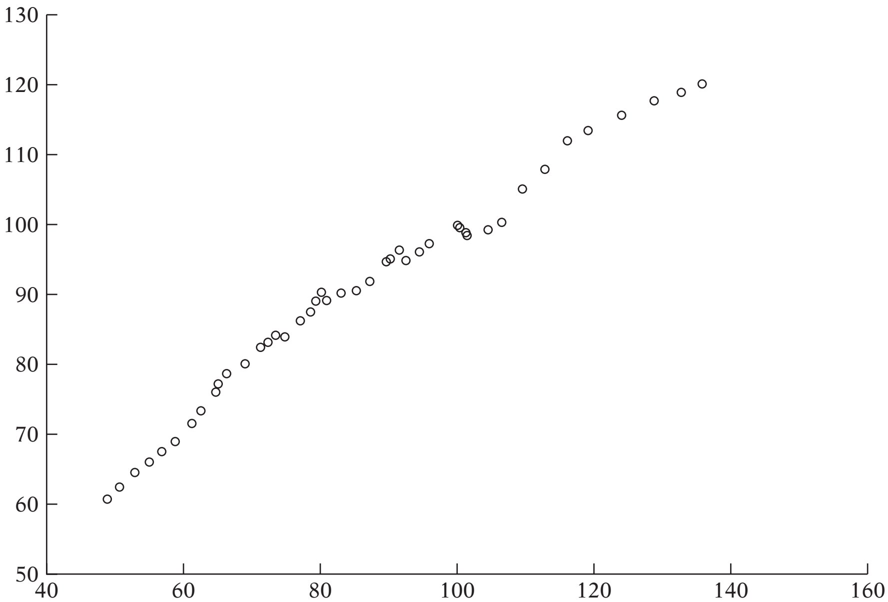
**TABLA 12.4**
**Índices de remuneración real y productividad en Estados Unidos, 1960-2005 (cifras de los índices, 1992 = 100; datos trimestrales ajustados por estacionalidad)**

Fuente: Economic Report of the President, 2007, tabla B-49.

| Año | Y | X | Año | Y | X |
|---|---|---|---|---|---|
| 1960 | 60.8 | 48.9 | 1983 | 90.3 | 83.0 |
| 1961 | 62.5 | 50.6 | 1984 | 90.7 | 85.2 |
| 1962 | 64.6 | 52.9 | 1985 | 92.0 | 87.1 |
| 1963 | 66.1 | 55.0 | 1986 | 94.9 | 89.7 |
| 1964 | 67.7 | 56.8 | 1987 | 95.2 | 90.1 |
| 1965 | 69.1 | 58.8 | 1988 | 96.5 | 91.5 |
| 1966 | 71.7 | 61.2 | 1989 | 95.0 | 92.4 |
| 1967 | 73.5 | 62.5 | 1990 | 96.2 | 94.4 |
| 1968 | 76.2 | 64.7 | 1991 | 97.4 | 95.9 |
| 1969 | 77.3 | 65.0 | 1992 | 100.0 | 100.0 |
| 1970 | 78.8 | 66.3 | 1993 | 99.7 | 100.4 |
| 1971 | 80.2 | 69.0 | 1994 | 99.0 | 101.3 |
| 1972 | 82.6 | 71.2 | 1995 | 98.7 | 101.5 |
| 1973 | 84.3 | 73.4 | 1996 | 99.4 | 104.5 |
| 1974 | 83.3 | 72.3 | 1997 | 100.5 | 106.5 |
| 1975 | 84.1 | 74.8 | 1998 | 105.2 | 109.5 |
| 1976 | 86.4 | 77.1 | 1999 | 108.0 | 112.8 |
| 1977 | 87.6 | 78.5 | 2000 | 112.0 | 116.1 |
| 1978 | 89.1 | 79.3 | 2001 | 113.5 | 119.1 |
| 1979 | 89.3 | 79.3 | 2002 | 115.7 | 124.0 |
| 1980 | 89.1 | 79.2 | 2003 | 117.7 | 128.7 |
| 1981 | 89.3 | 80.8 | 2004 | 119.0 | 132.7 |
| 1982 | 90.4 | 80.1 | 2005 | 120.2 | 135.7 |

*Notas:* $Y$ = índice de remuneración real por hora, sector de negocios (1992 = 100).
$X$ = índice de producción, sector de negocios (1992 = 100).

**FIGURA 12.7**
Índice de remuneración ($Y$) e índice de productividad ($X$), Estados Unidos, 1960-2005.




donde $d$ es el estadístico Durbin-Watson, que analizaremos en breve.

$$\begin{aligned} \widehat{\ln Y_t} &= \quad 1.6067 \;+\; 0.6522 \ln X_t \\ \text{ee} &= (0.0547) \quad (0.0124) \\ t &= (29.3680) \quad (52.7996) \\ r^2 &= 0.9845 \qquad d = 0.2176 \qquad \hat{\sigma} = 0.0221 \end{aligned} \tag{12.5.2}$$

En vista de que el modelo anterior es de doble logaritmo, el coeficiente de la pendiente representa la elasticidad. En el caso presente, observamos que si la productividad de la mano de obra aumenta 1%, la remuneración promedio se incrementa alrededor de 0.65%.

Cualitativamente, ambos modelos dan resultados semejantes. En los dos casos, los coeficientes estimados son "muy" significativos, como indican los altos valores $t$. En el modelo lineal, si el índice de productividad aumenta una unidad, en promedio, el de remuneración se incrementa casi 0.67 unidades. En el modelo log-lineal, al ser el coeficiente de pendiente una elasticidad (¿por qué?), tenemos que si el índice de productividad se incrementa 1%, en promedio, el índice de remuneración real aumenta casi 0.65%.

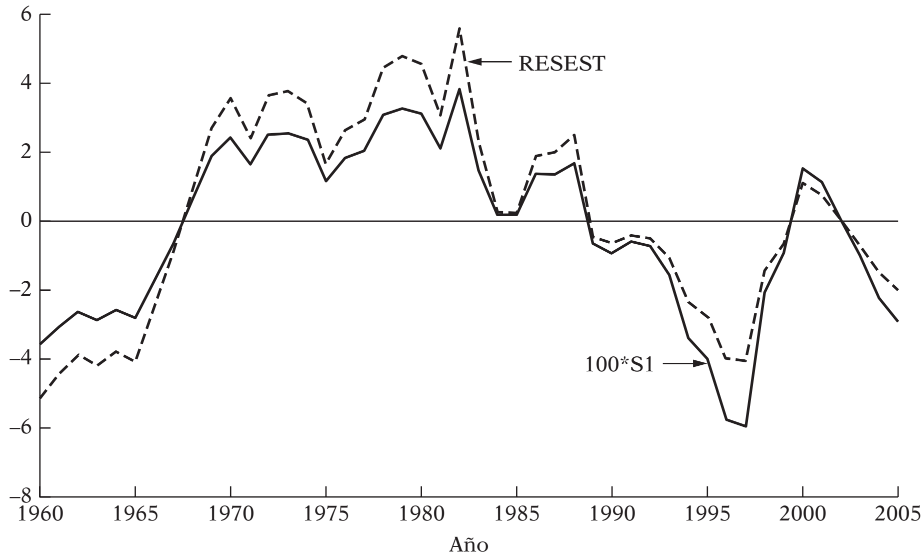
¿Cuán confiables son los resultados en (12.5.1) y (12.5.2) si existe autocorrelación? Como ya establecimos, si hay autocorrelación, los errores estándar estimados están sesgados, y como resultado, las razones $t$ estimadas no son confiables. Obviamente necesitamos averiguar si en los datos hay autocorrelación. En la siguiente sección analizamos varios métodos para detectar la autocorrelación. Ilustraremos estos métodos con el modelo log-lineal (12.5.2).

## 12.6 Detección de la autocorrelación

### I. Método gráfico

Recuerde que el supuesto de no autocorrelación del modelo clásico se relaciona con las perturbaciones poblacionales $u_t$, las cuales no pueden observarse directamente. En su lugar disponemos de valores sustitutos, los residuos $\hat{u}_t$, a partir del procedimiento usual MCO. Aunque las $\hat{u}_t$ no son

**FIGURA 12.8**
Residuos (amplificados 100 veces) y residuos estandarizados de la regresión de salarios sobre productividad (forma log: modelo 12.5.2).




lo mismo que las $u_t$,[17] con mucha frecuencia un examen visual de las $\hat{u}$ da algunas claves sobre la posible presencia de autocorrelación en las $u$. En realidad, un examen visual de $\hat{u}_t$ o ($\hat{u}_t^2$) proporciona información útil no sólo sobre la autocorrelación, sino también sobre heteroscedasticidad (como vimos en el capítulo anterior), sobre el grado de adecuación del modelo o sobre el sesgo de especificación, lo cual veremos en el siguiente capítulo. Como afirma un autor:

> No se puede exagerar la importancia de producir y analizar gráficos [de residuos] como parte habitual del análisis estadístico. Además de proporcionar en ocasiones un resumen accesible para entender un problema complejo, permiten el examen simultáneo de los datos, considerados en su conjunto, mientras que a la vez ilustran con claridad el comportamiento de los casos individuales.[18]

Hay diversas formas de examinar los residuos. Podemos graficarlos simplemente respecto del tiempo, con una **gráfica secuencial de tiempo**, como en la figura 12.8, que muestra los residuos obtenidos de la regresión de salarios sobre la productividad en Estados Unidos (12.5.2). Los valores de estos residuos están en la tabla 12.5, junto con algunos otros datos.

Por otro lado, podemos graficar los **residuos estandarizados** respecto del tiempo, los cuales también se muestran en la figura 12.8 y en la tabla 12.5. Los residuos estandarizados son tan sólo los residuos ($\hat{u}_t$) divididos entre el error estándar de la regresión ($\sqrt{\hat{\sigma}^2}$); es decir, son ($\hat{u}_t/\hat{\sigma}$). Observe que $\hat{u}_t$ al igual que $\hat{\sigma}$, están medidos en las unidades en las cuales se mide la variable regresada $Y$. Los valores de los residuos estandarizados serán números puros (desprovistos de unidades de medición) y, por consiguiente, son comparables con los residuos estandarizados de otras regresiones. Además, los residuos estandarizados, así como $\hat{u}_t$, tienen media igual a cero (¿por qué?) y varianza *aproximadamente* igual a la unidad.[19]

[17] Aunque las perturbaciones $u_t$ sean homoscedásticas y no estén correlacionadas, sus estimadores, los residuos $\hat{u}_t$, son heteroscedásticos y autocorrelacionados. Al respecto, véase G.S. Maddala, *Introduction to Econometrics*, Macmillan, 2a. ed., Nueva York, 1992, pp. 480-481. No obstante, se puede mostrar que, conforme se incrementa el tamaño de la muestra de manera indefinida, los residuos tienden a converger a sus valores reales, las $u_t$. Sobre este tema, consulte E. Malinvaud, *Statistical Methods of Econometrics*, 2a. ed., North-Holland Publishers, Ámsterdam, 1970, p. 88.

[18] Stanford Weisberg, *Applied Linear Regression*, John Wiley & Sons, Nueva York, 1980, p. 120.

[19] En realidad, son los llamados residuos de **Studentized** los que tienen varianza unitaria. Pero en la práctica, los residuos estandarizados suelen mostrar la misma condición y, por tanto, se puede confiar en ellos. Al respecto, véase Norman Draper y Harry Smith, *Applied Regression Analysis*, 3a. ed., John Wiley & Sons, Nueva York, 1998, pp. 207-208.

**TABLA 12.5** **Residuos: Reales, estandarizados y rezagados**

| Observación | S1 | RESEST | S1(−1) | Observación | S1 | RESEST | S1(−1) |
|---|---|---|---|---|---|---|---|
| 1960 | −0.036068 | −1.639433 | NA | 1983 | 0.014416 | 0.655291 | 0.038719 |
| 1961 | −0.030780 | −1.399078 | −0.036068 | 1984 | 0.001774 | 0.080626 | 0.014416 |
| 1962 | −0.026724 | −1.214729 | −0.030780 | 1985 | 0.001620 | 0.073640 | 0.001774 |
| 1963 | −0.029160 | −1.325472 | −0.026724 | 1986 | 0.013471 | 0.612317 | 0.001620 |
| 1964 | −0.026246 | −1.193017 | −0.029160 | 1987 | 0.013725 | 0.623875 | 0.013471 |
| 1965 | −0.028348 | −1.288551 | −0.026246 | 1988 | 0.017232 | 0.783269 | 0.013725 |
| 1966 | −0.017504 | −0.795647 | −0.028348 | 1989 | −0.004818 | −0.219005 | 0.017232 |
| 1967 | −0.006419 | −0.291762 | −0.017504 | 1990 | −0.006232 | −0.283285 | −0.004818 |
| 1968 | 0.007094 | 0.322459 | −0.006419 | 1991 | −0.004118 | −0.187161 | −0.006232 |
| 1969 | 0.018409 | 0.836791 | 0.007094 | 1992 | −0.005078 | −0.230822 | −0.004118 |
| 1970 | 0.024713 | 1.123311 | 0.018409 | 1993 | −0.010686 | −0.485739 | −0.005078 |
| 1971 | 0.016289 | 0.740413 | 0.024713 | 1994 | −0.023553 | −1.070573 | −0.010686 |
| 1972 | 0.025305 | 1.150208 | 0.016289 | 1995 | −0.027874 | −1.266997 | −0.023553 |
| 1973 | 0.025829 | 1.174049 | 0.025305 | 1996 | −0.039805 | −1.809304 | −0.027874 |
| 1974 | 0.023744 | 1.079278 | 0.025829 | 1997 | −0.041164 | −1.871079 | −0.039805 |
| 1975 | 0.011131 | 0.505948 | 0.023744 | 1998 | −0.013576 | −0.617112 | −0.041164 |
| 1976 | 0.018359 | 0.834515 | 0.011131 | 1999 | −0.006674 | −0.303364 | −0.013576 |
| 1977 | 0.020416 | 0.927990 | 0.018359 | 2000 | 0.010887 | 0.494846 | −0.006674 |
| 1978 | 0.030781 | 1.399135 | 0.020416 | 2001 | 0.007551 | 0.343250 | 0.010887 |
| 1979 | 0.033023 | 1.501051 | 0.030781 | 2002 | 0.000453 | 0.020599 | 0.007551 |
| 1980 | 0.031604 | 1.436543 | 0.033023 | 2003 | −0.006673 | −0.303298 | 0.000453 |
| 1981 | 0.020801 | 0.945516 | 0.031604 | 2004 | −0.015650 | −0.711380 | −0.006673 |
| 1982 | 0.038719 | 1.759960 | 0.020801 | 2005 | −0.020198 | −0.918070 | −0.015650 |

*Notas:* S1 = residuos de la regresión de salarios-productividad (forma log).
S1(−1) = residuos rezagados un periodo.
RESEST = residuos estandarizados = residuos/error estándar de la estimación.

En muestras grandes, $(\hat{u}_t/\hat{\sigma})$ está distribuida en forma aproximadamente normal con media cero y varianza unitaria. Para este ejemplo, $\hat{\sigma} = 2.6755$.

Al examinar la gráfica secuencial de tiempo de la figura 12.8, observamos que tanto $\hat{u}_t$ como $\hat{u}_t$ estandarizada presentan un patrón similar al de la figura 12.1*d*, lo que indica que tal vez las $u_t$ no sean aleatorias.

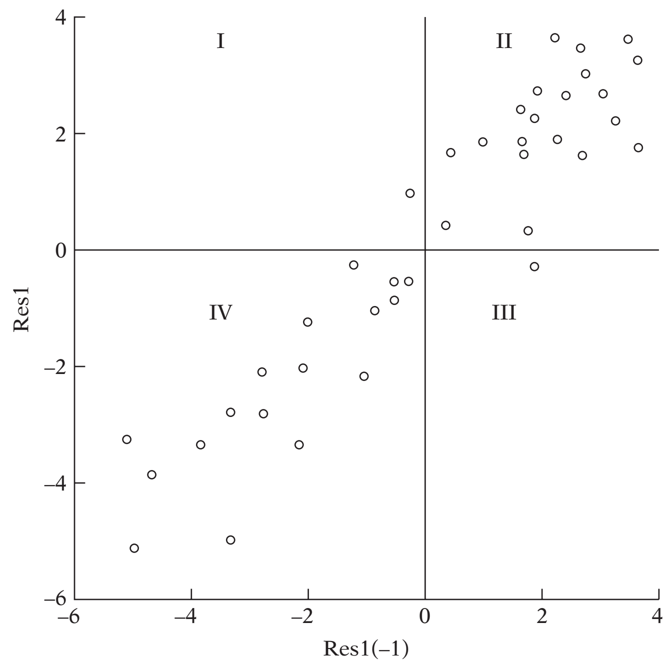
Para ver esto en forma diferente, podemos graficar $\hat{u}_t$ respecto de $\hat{u}_{t-1}$, es decir, el residuo en el tiempo $t$ frente a su valor en el tiempo $(t-1)$, una clase de prueba empírica del esquema AR(1). Si los residuos no son aleatorios, debemos obtener gráficas similares a las que aparecen en la figura 12.3. El gráfico de la regresión log de salarios-productividad se presenta en la figura 12.9; los datos básicos se proporcionan en la tabla 12.5. Como muestra esta figura, la mayoría de los residuos están agrupados en el segundo (noreste) y el cuarto (suroeste) cuadrantes, lo cual indica una correlación positiva fuerte en los residuos.

Por naturaleza, el método gráfico que acabamos de exponer es en esencia subjetivo o cualitativo, aunque poderoso. Sin embargo, hay diversas pruebas cuantitativas útiles para complementar el enfoque puramente cualitativo. A continuación veremos algunas de estas pruebas.

## II. Prueba de "las rachas"

Al examinar la figura 12.8, destaca una característica peculiar: al principio se tienen varios residuos negativos, luego se presenta una serie de residuos positivos y al final se observan muchos residuos, nuevamente negativos. Si los residuos fuesen puramente aleatorios, ¿sería posible ob-

**FIGURA 12.9**
Residuos actuales y residuos rezagados.




servar tal patrón? Por intuición, parece poco probable. Esta intuición se verifica con la llamada **prueba de "las rachas"**, conocida también como **prueba de Geary**, prueba no paramétrica.[20]

Para explicar esta prueba, se anotan simplemente los signos (+ o −) de los residuos obtenidos de la regresión salarios-productividad, que se presentan en la primera columna de la tabla 12.5.

$$(--------)(+++++++++++++++++++++)(-----------)(+++)(---) \tag{12.6.1}$$

Por tanto, hay 8 residuos negativos, seguidos por 21 positivos, seguidos por 11 negativos, seguidos por 3 positivos, seguidos por 3 negativos, para un total de 46 observaciones.

Definimos ahora una **racha** como una sucesión ininterrumpida de un símbolo o atributo, como + o −. Definimos además la **longitud de una racha** como el número de elementos que contiene. En la sucesión mostrada en (12.6.1), hay 5 rachas: una racha de 8 signos menos (es decir, de longitud 8), una racha de 21 signos más (es decir, de longitud 21), una racha de 11 signos menos (es decir, de longitud 11), una racha de 3 signos más (es decir, de longitud 3) y una racha de 3 signos menos (es decir, de longitud 3). Para un mejor efecto visual, presentamos las rachas entre paréntesis.

Al examinar el comportamiento de las rachas en una sucesión de observaciones estrictamente aleatoria, es posible derivar una prueba de aleatoriedad de las rachas. Nos planteamos la siguiente pregunta: ¿son muchas o muy pocas las 5 rachas observadas en el ejemplo ilustrativo consistente en 46 observaciones en comparación con el número de rachas esperadas en una sucesión de 46 observaciones estrictamente aleatoria? Si hay muchas rachas, significa que en el

[20] En las **pruebas no paramétricas** no se hacen supuestos sobre la distribución (de probabilidad) de dónde se obtuvieron las observaciones. Sobre la prueba de Geary, véase R.C. Geary, "Relative Efficiency of Count Sign Changes for Assessing Residual Autoregression in Least Squares Regression", *Biometrika*, vol. 57, 1970, pp. 123-127.

ejemplo los residuos cambian de signo frecuentemente, y se indica con esto una correlación serial negativa (compare esto con la figura 12.3*b*). En forma similar, si hay muy pocas rachas, pueden indicar autocorrelación positiva, como en la figura 12.3*a*). Entonces, *a priori*, la figura 12.8 indicaría una correlación positiva en los residuos.

Ahora, sea:

$N$ = número total de observaciones $= N_1 + N_2$

$N_1$ = número de símbolos + (es decir, residuos +)

$N_2$ = número de símbolos − (es decir, residuos −)

$R$ = número de rachas

Entonces, según la hipótesis nula de que los resultados sucesivos (en este caso, residuos) son independientes, y si suponemos que $N_1 > 10$ y $N_2 > 10$, el número de rachas está (*asintóticamente*) *normalmente distribuido* con

$$\text{Media:} \qquad E(R) = \frac{2N_1N_2}{N} + 1$$

$$\text{Varianza:} \qquad \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{(N)^2(N-1)} \tag{12.6.2}$$

*Nota:* $N = N_1 + N_2$.

Si la hipótesis nula de aleatoriedad es sostenible, y según las propiedades de la distribución normal, debemos esperar que

$$\text{Prob}\,[E(R) - 1.96\sigma_R \leq R \leq E(R) + 1.96\sigma_R] = 0.95 \tag{12.6.3}$$

Es decir, la probabilidad de que el intervalo anterior incluya a $R$ es de 95%. Por tanto, tenemos la siguiente regla:

***Regla de decisión***

No rechace la hipótesis nula de aleatoriedad a 95% de confianza si *R*, el número de rachas, está en el intervalo de confianza anterior; rechace la hipótesis nula si la *R* estimada se encuentra fuera de estos límites. (*Nota:* Puede elegir cualquier nivel de confianza que desee.)

Regresemos al ejemplo, donde sabemos que $N_1$, el número de signos positivos, es 24, y $N_2$, el número de signos negativos, es 22; además $R = 5$. Con las fórmulas dadas en (12.6.2), obtenemos:

$$E(R) = 24$$
$$\sigma_R^2 = 11 \tag{12.6.4}$$
$$\sigma_R = 3.32$$

El intervalo de confianza a 95% para $R$ en el ejemplo es entonces:

$$[24 \pm 1.96(3.32)] = (17.5, 30.5)$$

Obvio, este intervalo no incluye 5. Por tanto, *rechazamos* la hipótesis de que los residuos en la regresión de los salarios sobre la productividad son aleatorios, con una confianza de 95%. En otras palabras, los residuos muestran autocorrelación. Como regla general, si hay autocorrelación positiva, el número de rachas será reducido, mientras que si existe autocorrelación negativa, el

número de rachas será grande. Por supuesto, de (12.6.2) sabemos si se tienen muchas o pocas rachas.

Swed y Eisenhart elaboraron tablas especiales con valores críticos de las rachas esperadas en una sucesión aleatoria de $N$ observaciones, si $N_1$ o $N_2$ son menores que 20. Estas tablas se incluyen en el **apéndice D**, tabla D.6. Con esas tablas, el lector puede verificar que los residuos en la regresión de salarios sobre productividad en realidad no son aleatorios; de hecho, están positivamente correlacionados.

## III. Prueba *d* de Durbin-Watson[21]

La prueba más conocida para detectar correlación serial es la de los estadísticos Durbin y Watson. Se le conoce como **estadístico *d* de Durbin-Watson**, que se define como

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2} \tag{12.6.5}$$

que es simplemente la razón de la suma de las diferencias al cuadrado de residuos sucesivos sobre la SCR. Observe que, en el numerador del estadístico $d$, el número de observaciones es $n - 1$ porque se pierde una observación al obtener las diferencias consecutivas.

Una gran ventaja del estadístico $d$ es que se basa en los residuos estimados, que se calculan de manera rutinaria en los análisis de regresión. Debido a esta ventaja, es frecuente incluir el estadístico $d$ de Durbin-Watson en los informes de análisis de regresión, junto con otros estadísticos de resumen, como $R^2$, $R^2$ ajustada, $t$ y $F$. Aunque el estadístico $d$ se utiliza ahora en forma rutinaria, **es importante observar los supuestos en los cuales se basa**:

1. El modelo de regresión incluye el término del intercepto. Si dicho término no está presente, como en la regresión a través del origen, es esencial efectuar de nuevo la regresión con dicho término para obtener la SCR.[22]

2. Las variables explicativas, $X$, son no estocásticas, es decir, son fijas en muestreo repetido.

3. Las perturbaciones $u_t$ se generan mediante el esquema autorregresivo de primer orden: $u_t = \rho u_{t-1} + \varepsilon_t$. Por tanto, no se pueden utilizar para detectar esquemas autorregresivos de orden superior.

4. Se supone que el término de error $u_t$ está normalmente distribuido.

5. El modelo de regresión no incluye valor(es) rezagado(s) de la variable dependiente como una variable explicativa. Por tanto, la prueba es inaplicable a modelos del siguiente tipo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t \tag{12.6.6}$$

donde $Y_{t-1}$ es el valor de $Y$ rezagada un periodo. Tales modelos se conocen como **modelos autorregresivos**, los cuales estudiaremos en el capítulo 17.

6. No hay observaciones faltantes en los datos. Por tanto, en la regresión de salarios-productividad de 1960 a 2005, si por alguna razón faltaran observaciones, por ejemplo, de 1978 y 1982, el estadístico $d$ no permitiría la ausencia de tales observaciones.[23]

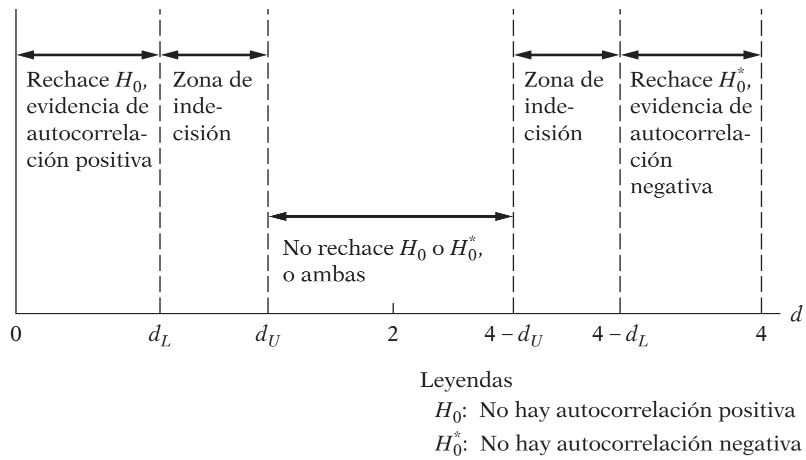
---

[21] J. Durbin y G.S. Watson, "Testing for Serial Correlation in Least-Squares Regression", *Biometrika*, vol. 38, 1951, pp. 159-171.

[22] Sin embargo, R.W. Farebrother calculó valores *d* cuando el término del intercepto está ausente del modelo. Véase su artículo "The Durbin-Watson Test for Serial Correlation When There Is No Intercept in the Regression", *Econometrica*, vol. 48, 1980, pp. 1553-1563.

[23] Para mayores detalles, véase Gabor Korosi, Laszlo Matyas e Istvan P. Szekey, *Practical Econometrics*, Avebury Press, Inglaterra, 1992, pp. 88-89.

**FIGURA 12.10**
Estadístico *d* de Durbin-Watson.




El muestreo exacto o la distribución de probabilidad del estadístico $d$ dado en (12.6.5) es difícil de derivar porque, como demostraron Durbin y Watson, tiene una dependencia compleja de los valores presentes de $X$ en una muestra dada.[24] Esta dificultad se entiende porque $d$ se calculó a partir de los $\hat{u}_t$, los cuales, por supuesto, dependen de las $X$ dadas. Por consiguiente, a diferencia de las pruebas $t$, $F$ o $\chi^2$, no hay un valor crítico único que lleve al rechazo o a la aceptación de la hipótesis nula de que no hay correlación serial de primer orden en las perturbaciones $u_i$. Sin embargo, Durbin y Watson lograron encontrar un límite inferior $d_L$ y un límite superior $d_U$ tales que si el valor $d$ calculado de (12.6.5) cae por fuera de estos valores críticos, puede tomarse una decisión respecto de la presencia de correlación serial positiva o negativa. Además, estos límites sólo dependen del número de observaciones $n$ y del número de variables explicativas, y no de los valores que adquieren estas variables explicativas. Durbin y Watson tabularon estos límites para $n$, de 6 a 200 y hasta 20 variables explicativas, y se presentan en el **apéndice D**, tabla D.5 (hasta 20 variables explicativas).

El procedimiento de prueba aplicado se explica mejor con ayuda de la figura 12.10, la cual muestra que los límites de $d$ son 0 y 4. Éstos se determinan al expandir (12.6.5) para obtener

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.7}$$

Como $\sum \hat{u}_t^2$ y $\sum \hat{u}_{t-1}^2$ difieren sólo en una observación, son aproximadamente iguales. Por consiguiente, establecemos que $\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$ y (12.6.7) se escribe como

$$d \approx 2\left(1 - \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \tag{12.6.8}$$

donde $\approx$ significa aproximadamente igual.

Ahora definimos

$$\hat{\rho} = \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.9}$$

[24] Sin embargo, consulte el análisis sobre la prueba Durbin-Watson “exacta” más adelante en la sección.

como el coeficiente de autocorrelación muestral de primer orden, un estimador de $\rho$. (Véase la nota 9.) Con (12.6.9) expresamos (12.6.8) como

$$d \approx 2(1 - \hat{\rho}) \tag{12.6.10}$$

Pero, como $-1 \leq \rho \leq 1$ (12.6.10), implica que

$$0 \leq d \leq 4 \tag{12.6.11}$$

Estos son los límites de $d$; todo valor $d$ estimado debe caer dentro de estos límites.

Es evidente de la ecuación (12.6.10) que si $\hat{\rho} = 0$, $d = 2$; es decir, si no hay correlación serial (de primer orden), esperamos que $d$ esté alrededor de 2. *Por consiguiente, como regla práctica, si en una aplicación vemos que* d *es igual a 2, podemos suponer que no hay autocorrelación de primer orden, positiva o negativa*. Si $\hat{\rho} = +1$, indica una correlación positiva perfecta en los residuos, $d \approx 0$. Por consiguiente, entre más cercano esté $d$ a 0, mayor será la evidencia de correlación serial positiva. Esta relación debe ser evidente de (12.6.5) porque, si hay autocorrelación positiva, las $\hat{u}_t$ aparecerán agrupadas, y sus diferencias, por consiguiente, tenderán a ser pequeñas. Como resultado, la suma de cuadrados del numerador será menor en comparación con la suma de cuadrados del denominador, el cual es un valor que permanece fijo para cualquier regresión dada.

Si $\hat{\rho} = -1$ es decir, hay una correlación negativa perfecta entre los valores consecutivos de los residuos, $d \approx 4$. Por tanto, entre más se acerque $d$ a 4, mayor será la evidencia de correlación serial negativa. De nuevo, al analizar (12.6.5), esto es comprensible: si hay autocorrelación negativa, una $\hat{u}_t$ positiva tenderá a estar seguida por una $\hat{u}_t$ negativa y viceversa, de forma que $|\hat{u}_t - \hat{u}_{t-1}|$ será usualmente mayor que $|\hat{u}_t|$. Por consiguiente, el numerador de $d$ será comparativamente mayor que el denominador.

El mecanismo de la prueba de Durbin-Watson es el siguiente, si suponemos que se cumplen los supuestos de la prueba:

1. Efectuar la regresión por MCO y obtener los residuos.
2. Calcular $d$ a partir de (12.6.5). (La mayoría de los programas de computadora incluye este cálculo.)
3. Para un tamaño de muestra dado y un número de variables explicativas dado, determinar los valores críticos $d_L$ y $d_U$.
4. Ahora se siguen las reglas de decisión de la tabla 12.6. Para facilitar su comprensión, estas reglas se resumen en la figura 12.10.

Para ilustrar el mecanismo, retome la regresión salarios-productividad. De los datos en la tabla 12.5 vemos que el valor estimado de $d$ es 0.2175, lo que indica una correlación serial positiva en los residuos. A partir de las tablas de Durbin-Watson, encontramos que, para 46 observaciones y una variable explicativa, $d_L = 1.475$ y $d_U = 1.566$ en el nivel de 5%. Como el valor $d$ calculado de 0.2175 está por debajo de $d_L$, no podemos rechazar la hipótesis de que hay correlación serial positiva en los residuos.

A pesar de ser muy popular, la prueba $d$ tiene una gran desventaja: cuando cae en la **zona de indecisión**, no se puede concluir si hay o no autocorrelación (de primer orden). Para resolver

**TABLA 12.6**
**Prueba *d* de Durbin-Watson: reglas de decisión**

| Hipótesis nula | Decisión | Si |
|---|---|---|
| No hay autocorrelación positiva | Rechazar | $0 < d < d_L$ |
| No hay autocorrelación positiva | Sin decisión | $d_L \leq d \leq d_U$ |
| No hay correlación negativa | Rechazar | $4 - d_L < d < 4$ |
| No hay correlación negativa | Sin decisión | $4 - d_U \leq d \leq 4 - d_L$ |
| No hay autocorrelación, positiva o negativa | No rechazar | $d_U < d < 4 - d_U$ |

este problema, diversos autores proponen modificaciones a la prueba $d$, las cuales escapan al alcance de este libro.[25] En muchas situaciones, sin embargo, se observa que el límite superior $d_U$ es aproximadamente el verdadero límite de significancia, y, por consiguiente, en el caso de que el valor $d$ estimado se encuentre en la zona de indecisión, se puede utilizar el siguiente procedimiento de **prueba *d* modificada**. Con el nivel de significancia $\alpha$,

1. $H_0{:}\rho = 0$ frente a $H_1{:}\rho > 0$. Si el valor estimado $d < d_U$, rechace $H_0$ en el nivel $\alpha$. Es decir, hay correlación positiva estadísticamente significativa.
2. $H_0{:}\rho = 0$ frente a $H_1{:}\rho < 0$. Si el valor estimado $(4 - d) < d_U$, rechace $H_0$ en el nivel $\alpha$; es decir, hay evidencia estadísticamente significativa de autocorrelación negativa.
3. $H_0{:}\rho = 0$ frente a $H_1{:}\rho \neq 0$. Rechace $H_0$ en el nivel $2\alpha$ si $d < d_U$ o $(4 - d) < d_U$, es decir, hay evidencia estadísticamente significativa de autocorrelación, positiva o negativa.

Se debe señalar que la zona de indecisión se estrecha conforme aumenta el tamaño de la muestra, lo cual se observa claramente en las tablas Durbin-Watson. Por ejemplo, con 4 regresoras y 20 observaciones, los valores $d$ inferiores y superiores a 5% son 0.894 y 1.828, respectivamente; pero esos valores son 1.515 y 1.739, cuando el tamaño de la muestra es de 75.

El software SHAZAM calcula una *prueba d exacta*; es decir, da el valor $p$, la probabilidad exacta del valor $d$ calculado. Con las computadoras modernas, ya no es difícil determinar el valor $p$ del estadístico $d$ calculado. Con SHAZAM (versión 9) para la regresión de salarios sobre productividad, tenemos que el valor $p$ del $d$ calculado es 0.2176, prácticamente cero, por lo que confirmamos la conclusión anterior basada en las tablas Durbin-Watson.

La prueba $d$ de Durbin-Watson es ya tan clásica que los profesionales suelen olvidar los supuestos en los que se basa; en particular: 1) las variables explicativas, o regresoras, son no estocásticas, 2) el término de error sigue la distribución normal, 3) los modelos de regresión no incluyen el (los) valor(es) rezagado(s) de la regresada y 4) sólo se toma en cuenta la correlación serial de primer orden. Todos estos supuestos son muy importantes en la aplicación de la prueba $d$. Debe añadirse que un estadístico $d$ significativo no necesariamente indica autocorrelación. Más bien, puede indicar una omisión de las variables pertinentes en el modelo.

Si un modelo de regresión contiene valor(es) rezagado(s) de la regresada, el valor $d$ a menudo se aproxima a 2, lo cual indicaría que no hay autocorrelación (de primer orden) en dichos modelos. Por tanto, hay un sesgo implícito que impide descubrir la autocorrelación (de primer orden) en tales modelos. Esto no quiere decir que los modelos autorregresivos estén exentos de problemas de autocorrelación. De hecho, Durbin elaboró la llamada **prueba *h*** para probar en tales modelos la correlación serial; sin embargo, no es tan poderosa, en sentido estadístico, como la **prueba Breusch-Godfrey**, que en breve analizaremos, así que no hay necesidad de la **prueba *h***. Sin embargo, por su importancia histórica, la abordaremos en el ejercicio 12.36.

Asimismo, si los términos de error $u_t$ no son NIID, tal vez no sea confiable la prueba $d$ rutinaria.[26] Al respecto, la **prueba de las rachas** tiene la ventaja de que no hace ningún supuesto en cuanto a la distribución (de probabilidad) respecto del término de error. Sin embargo, si la muestra es grande (técnicamente infinita) se puede utilizar el $d$ de Durbin-Watson, pues se demuestra que:[27]

$$\sqrt{n}\left(1 - \frac{1}{2}d\right) \approx N(0, 1) \qquad \textbf{(12.6.12)}$$

[25] Para mayores detalles, véase Thomas B. Fomby, R. Carter Hill y Stanley R. Johnson, *Advanced Econometric Methods*, Springer-Verlag, Nueva York, 1984, pp. 225-228.

[26] Para un análisis avanzado, véase Ron C. Mittelhammer, George G. Judge y Douglas J. Millar, *Econometric Foundations*, Cambridge University Press, Nueva York, 2000, p. 550.

[27] Véase James Davidson, *Econometric Theory*, Blackwell, Nueva York, 2000, p. 161.

Es decir, en muestras grandes, el estadístico $d$, como está transformado en (12.6.12), sigue la distribución normal estandarizada. A propósito, en vista de la relación entre $d$ y $\hat{\rho}$, el coeficiente estimado de autocorrelación de primer orden, mostrado en (12.6.10), se deduce que

$$\sqrt{n}\hat{\rho} \approx N(0, 1) \tag{12.6.13}$$

es decir, en muestras grandes, la raíz cuadrada del tamaño de la muestra multiplicada por el coeficiente de autocorrelación de primer orden también sigue la distribución normal estandarizada.

Como una ilustración de esta prueba, para el ejemplo de la regresión de los salarios sobre la productividad descubrimos que $d = 0.2176$, con $n = 46$. Por tanto, de (12.6.12) tenemos que

$$\sqrt{46}\left(1 - \frac{0.2176}{2}\right) \approx 6.0447$$

Asintóticamente, si fuese verdadera la hipótesis nula de que no existe autocorrelación (de primer orden), la probabilidad de obtener un valor $Z$ (es decir, la variable normal estandarizada) igual a 6.0447 o mayor sería extremadamente pequeña. Recuerde que para una distribución normal estandarizada, el valor crítico $Z$ a 5% (de dos colas) tan sólo es de 1.96, y el valor crítico $Z$ a 1% es de casi 2.58. Aunque el tamaño de la muestra es únicamente de 46, para propósitos prácticos puede ser lo bastante grande para utilizar la aproximación normal. La conclusión es aún la misma, a saber, que los residuos de la regresión de los salarios sobre la productividad presentan autocorrelación.

Pero el problema más grave con la prueba $d$ es el supuesto de que las regresoras son no estocásticas; es decir, sus valores son fijos en muestras repetidas. De no ser así, la prueba $d$ no es válida para muestras finitas, o pequeñas, ni para muestras grandes.[28] Y en vista de que este supuesto a menudo resulta difícil de sostener en los modelos económicos con datos de series de tiempo, un autor afirma que el estadístico Durbin-Watson tal vez no sea útil en la econometría con series de tiempo.[29] Su opinión es que hay pruebas más útiles para la autocorrelación, pero se basan en muestras grandes. Analizaremos a continuación una de estas pruebas, la **prueba de Breusch-Godfrey**.

## IV. Una prueba general de autocorrelación: la prueba de Breusch-Godfrey (BF)[30]

Para evitar algunos inconvenientes de la prueba $d$ de Durbin-Watson de autocorrelación, los estadísticos Breusch y Godfrey elaboraron una prueba para la autocorrelación que es general porque permite: 1) regresoras no estocásticas, como los valores rezagados de la regresada; 2) esquemas autorregresivos de orden mayor, como el AR(1), AR(2), etc.; y 3) **promedios móviles simples o de orden superior** de los términos de error de ruido blanco, como $\varepsilon_t$ en (12.2.1).[31]

Sin abordar los detalles matemáticos, los cuales puede consultar en la bibliografía, la **prueba BG**, que también se conoce como **prueba ML**,[32] procede de la siguiente manera: utilizamos el

[28] *Ibid.*, p. 161.

[29] Fumio Hayashi, *Econometrics*, Princeton University Press, Princeton, Nueva Jersey, 2000, p. 45.

[30] Véase L.G. Godfrey, "Testing Against General Autoregressive and Moving Average Error Models When the Regressor include Lagged Dependent Variable", *Econometrica*, vol. 46, 1978, pp. 1293-1302, y T.S. Breusch, "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economic Papers*, vol. 17, 1978, pp. 334-355.

[31] Por ejemplo, en la regresión $Y_t = \beta_1 + \beta_2 X_t + u_t$ el término de error puede expresarse por medio de $u_t = \varepsilon_t + \lambda_1\varepsilon_{t-1} + \lambda_2\varepsilon_{t-2}$, que representa un promedio móvil de tres periodos del término de error de ruido blanco $\varepsilon_t$.

[32] La prueba se basa en el **principio multiplicador de Lagrange**, mencionado brevemente en el capítulo 8.

modelo de regresión de dos variables para ilustrar la prueba, aunque se pueden añadir al modelo muchas regresoras. Asimismo, se pueden incluir en él valores rezagados de la regresada. Sea

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.6.14}$$

Suponga que el término de error $u_t$ sigue el esquema autorregresivo de orden $p$, AR($p$), del siguiente modo:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t \tag{12.6.15}$$

donde $\varepsilon_t$ es un término de error de ruido blanco, como ya examinamos. Lo anterior es una simple extensión del esquema AR(1), como el lector ya habrá reconocido.

La hipótesis nula $H_0$ por demostrar es

$$H_0\colon \rho_1 = \rho_2 = \cdots = \rho_p = 0 \tag{12.6.16}$$

Es decir, no existe correlación serial de ningún orden. La prueba BG implica los siguientes pasos:

1. Estime (12.6.14) mediante MCO y obtenga los residuos $\hat{u}_t$.

2. Haga la regresión $\hat{u}_t$ sobre la $X_t$ original (si hay más de una variable $X$ en el modelo original, inclúyalas también) y $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}$, donde estas últimas son los valores rezagados de los residuos estimados en el paso 1. Por tanto, si $p = 4$, introduciremos en el modelo cuatro valores rezagados de los residuos como regresoras adicionales. Observe que para hacer esta regresión sólo hay $(n - p)$ observaciones (¿por qué?). En resumen, realice la siguiente regresión:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t \tag{12.6.17}$$

y obtenga $R^2$ de esta regresión (auxiliar).[33]

3. Si el tamaño de la muestra es grande, Breusch y Godfrey demostraron que

$$(n - p)R^2 \sim \chi^2_p \tag{12.6.18}$$

Es decir, asintóticamente, $n - p$ veces el valor de $R^2$ obtenido en la regresión auxiliar (12.6.17) sigue la distribución ji cuadrada con $p$ gl. Si en una aplicación $(n - p)R^2$ excede el valor crítico ji cuadrada en el nivel de significancia seleccionado, podemos rechazar la hipótesis nula, en cuyo caso, por lo menos una $\rho$ en (12.6.15) es significativamente diferente de cero.

Pueden mencionarse los siguientes *puntos prácticos* sobre la prueba BG:

1. Las regresoras incluidas en el modelo de regresión pueden contener valores rezagados de la variable regresada $Y$; es decir, $Y_{t-1}$, $Y_{t-2}$, etc., pueden aparecer como variables explicativas. Contraste este modelo con la restricción de la prueba de Durbin-Watson, que no permite valores rezagados de la variable regresada entre las variables explicativas.

2. Como ya señalamos, la prueba BG es aplicable aunque las perturbaciones sigan un proceso de **promedios móviles (PM)** de orden $p$, es decir, aunque las $u_i$ se generen como sigue:

$$u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \cdots + \lambda_p \varepsilon_{t-p} \tag{12.6.19}$$

donde $\varepsilon_t$ es un término de error de ruido blanco; es decir, el término de error que satisface todos los supuestos clásicos.

---

[33] La razón para incluir la regresora original *X* en el modelo es permitir que *X* no sea estrictamente no estocástica. Pero si es estrictamente no estocástica, quizá se omita del modelo. Sobre este tema, consulte Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, South-Western Publishing Co., 2003, p. 386.

En los capítulos sobre la econometría de series de tiempo estudiaremos con cierto detalle los procesos de promedio móvil y los autorregresivos de orden $p$.

3. Si $p = 1$ en (12.6.15), que significa autorregresión de primer orden, la prueba BG se conoce como **prueba *m* de Durbin**.

4. Una desventaja de la prueba BG es que el valor de $p$, la longitud del rezago, no puede especificarse *a priori*. Es inevitable algún grado de experimentación con el valor de $p$. A veces se pueden utilizar los llamados criterios de información **Akaike** y **Schwarz** para seleccionar la longitud del rezago. Estudiaremos estos criterios en el capítulo 13 y más adelante, en los capítulos sobre econometría de series de tiempo.

5. Con los valores de las variables $X$ y los valores rezagados de $u$, la prueba supone que la varianza de $u$ en la ecuación (12.6.15) es homoscedástica.

*Ilustración de la prueba BG: Relación entre salarios y productividad*

Para mostrar la prueba, la aplicaremos a este ejemplo ilustrativo. Con un esquema AR(6) obtuvimos los resultados del ejercicio 12.25. De los resultados de la regresión dados ahí, se observa que $(n - p) = 40$ y $R^2 = 0.7498$. Por tanto, al multiplicar estos dos valores se obtiene un valor ji cuadrada de 29.992. Para 6 gl (¿por qué?), la probabilidad de obtener un valor ji cuadrada igual o mayor que 29.992 es demasiado pequeña; la tabla ji cuadrada del apéndice D.4 muestra que la probabilidad de obtener un valor ji cuadrada de 18.5476 o mayor es de sólo 0.005. Por tanto, para los mismos gl, la probabilidad de obtener un valor ji cuadrada de casi 30 debe ser demasiado pequeña. De hecho, el valor $p$ real es casi cero.

En consecuencia, la conclusión es que, para este ejemplo, al menos una de las seis autocorrelaciones debe ser distinta de cero.

Al tratar de variar las longitudes del rezago de 1 a 6, encontramos que sólo el coeficiente AR(1) es significativo, lo cual sugiere que no hay necesidad de considerar más que un rezago. En esencia, la prueba BG, en este caso, resulta ser la **prueba *m* de Durbin.**

### ¿Por qué tantas pruebas para la autocorrelación?

La respuesta es que ". . . no se ha juzgado de manera inequívoca a ninguna prueba en particular como la mejor [es decir, la más poderosa en el sentido estadístico], y por tanto el analista todavía está en la nada envidiable posición de considerar una variada colección de procedimientos de prueba para detectar la presencia, la estructura, o ambas, de la autocorrelación".[34] Por supuesto, un argumento similar vale respecto de las diversas pruebas de heteroscedasticidad analizadas en el capítulo anterior.

## 12.7 Qué hacer cuando hay autocorrelación: medidas correctivas

Si después de aplicar una o más pruebas de diagnóstico para la autocorrelación de las analizadas en la sección previa encontramos autocorrelación, ¿qué hacer? Hay cuatro opciones:

1. Trate de averiguar si se trata de **autocorrelación pura** y no el resultado de una mala especificación del modelo. Como analizamos en la sección 12.1, a veces se observan patrones en los residuos porque el modelo está mal especificado —es decir, se excluyeron variables importantes— o porque su forma funcional no es correcta.

[34] Ron C. Mittelhammer *et al.*, *op. cit.*, p. 547. Recuerde que la **potencia de una prueba estadística** es 1 menos la probabilidad de cometer un error de tipo II; es decir, 1 menos la probabilidad de aceptar una hipótesis falsa. La potencia máxima de una prueba es 1, y la mínima es 0. Mientras más cerca de cero esté la potencia de una prueba, peor será ésta, y mientras más cerca esté de 1, más poderosa será. Lo que estos autores afirman en esencia es que no existe una prueba de autocorrelación que sea la más poderosa.

2. Si se trata de autocorrelación pura, se puede utilizar una transformación apropiada del modelo original de manera que en el modelo transformado no se presente el problema de la autocorrelación (pura). Como en la heteroscedasticidad, habrá que emplear algún **método generalizado de mínimos cuadrados (MCG)**.

3. En muestras grandes se puede utilizar el método **Newey-West** para obtener los errores estándar de los estimadores de MCO corregidos para autocorrelación. Este método en realidad es una extensión del método de errores estándar consistentes con heteroscedasticidad de White, que analizamos en el capítulo anterior.

4. En algunas situaciones se puede conservar el método MCO.

Debido a la importancia de cada uno de estos temas, les dedicamos una sección.

## 12.8 Especificación incorrecta del modelo frente a autocorrelación pura

Consideremos de nuevo la regresión de salarios sobre productividad dada en (12.5.2). Vimos que el valor $d$ era igual a 0.2176, y con base en la prueba $d$ de Durbin-Watson concluimos que existía una correlación positiva en el término de error. ¿Pudo surgir dicha correlación porque el modelo no estaba correctamente especificado? Como los datos subyacentes a la regresión (12.5.1) son del tipo serie de tiempo, es muy probable que los salarios y la productividad muestren tendencias. De ser así, se requiere incluir la variable de tiempo o tendencia, $t$, en el modelo, a fin de observar la relación entre los salarios y la productividad descontando la tendencia de las dos variables.

Para probar lo anterior, incluimos la variable tendencia en (12.5.2) y obtuvimos los siguientes resultados:

$$\begin{aligned} \hat{Y}_t &= \ 0.1209 \ + \ 1.0283X_t - \ 0.0075t \\ \text{ee} &= (0.3070) \quad (0.0776) \quad (0.0015) \\ t &= (0.3939) \quad (13.2594) \quad (-4.8903) \\ & R^2 = 0.9900; \qquad d = 0.4497 \end{aligned} \tag{12.8.1}$$

La interpretación de este modelo es sencilla: con el tiempo, el índice de los salarios reales disminuyó casi 0.75 unidades al año. Después de tener esto en cuenta, si el índice de productividad aumentara una unidad, en promedio, el salario real se incrementaría casi una unidad. Resulta interesante notar que incluso teniendo en cuenta la variable de tendencia, el valor $d$ sigue siendo muy bajo, lo cual indica que (12.8.1) presenta autocorrelación pura y no necesariamente hay un error de especificación.

¿Cómo sabemos que (12.8.1) es la especificación correcta? Para averiguarlo, hacemos la regresión de $Y$ sobre $X$ y $X^2$ para probar la posibilidad de que el índice de salarios reales esté relacionado de forma no lineal con el índice de productividad. Los resultados de esta regresión son los siguientes:

$$\begin{aligned} \hat{Y}_t &= -\ 1.7843 \ + \ 2.1963X_t - \ 0.1752X_t^2 \\ t &= (-2.7713) \quad (7.5040) \quad (-5.2785) \\ & R^2 = 0.9906 \qquad d = 0.3561 \end{aligned} \tag{12.8.2}$$

Corresponde al lector interpretar estos resultados. Para los propósitos presentes, examine el valor Durbin-Watson, aún muy bajo, lo que indica que todavía hay correlación serial positiva en los residuos.

Se puede concluir con toda seguridad, a partir del análisis anterior, que la regresión de los salarios sobre la productividad presenta autocorrelación pura, y no necesariamente un sesgo de

especificación. Al conocer las consecuencias de la autocorrelación, quizá desearíamos emprender algunas acciones correctivas, lo cual haremos en breve.

A propósito, en todas las regresiones de salarios sobre productividad que se han presentado, se aplicó la **prueba de normalidad de Jarque-Bera** y se encontró que los residuos estaban normalmente distribuidos, lo cual resulta reconfortante porque la prueba *d* supone la normalidad para el término de error.

## 12.9 Corrección de la autocorrelación (pura): el método de los mínimos cuadrados generalizados (MCG)

Como conocemos las consecuencias de la autocorrelación, sobre todo la falta de eficiencia de los estimadores de MCO, quizá deseemos corregir el problema. El remedio depende del conocimiento respecto a la naturaleza de la interdependencia entre las perturbaciones; es decir, conocer la estructura de la autocorrelación.

Para empezar, considere el modelo de regresión de dos variables:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.9.1}$$

y suponga que el término de error sigue el esquema AR(1), a saber:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \tag{12.9.2}$$

Ahora consideremos dos casos: 1) se conoce $\rho$ y 2) no se conoce $\rho$, pero tiene que estimarse.

### Cuando se conoce $\rho$

Si se conoce el coeficiente de autocorrelación de primer orden, el problema de la autocorrelación se resuelve muy fácil. Si (12.9.1) es válida en el tiempo $t$, también lo es para el tiempo $(t-1)$. Por tanto,

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \tag{12.9.3}$$

Al multiplicar (12.9.3) por $\rho$ en ambos miembros, obtenemos

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \tag{12.9.4}$$

Si restamos (12.9.4) de (12.9.1), resulta

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \tag{12.9.5}$$

donde $\varepsilon_t = (u_t - \rho u_{t-1})$.

Expresamos (12.9.5) como

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \tag{12.9.6}$$

donde $\beta_1^* = \beta_1(1-\rho)$, $Y_t^* = (Y_t - \rho Y_{t-1})$, $X_t^* = (X_t - \rho X_{t-1})$ y $\beta_2^* = \beta_2$.

Como el término de error en (12.9.6) satisface los supuestos usuales de MCO, los aplicamos a las variables transformadas $Y^*$ y $X^*$ para obtener estimadores con todas las propiedades óptimas; a saber, que sean MELI. En efecto, llevar a cabo la regresión (12.9.6) equivale a utilizar los mínimos cuadrados generalizados (MCG) del capítulo anterior (recuerde que los MCG no son más que los MCO aplicados al modelo transformado que satisface los supuestos clásicos).

La regresión (12.9.5) se conoce como **generalizada**, **cuasi generalizada** o **ecuación en diferencias**. Implica la regresión de $Y$ sobre $X$, pero no en la forma original, sino en la **forma de diferencias**, que se obtiene al restar una proporción $(= \rho)$ del valor de una variable en el periodo

anterior, de su valor en el periodo actual. En este proceso de diferenciación se pierde una observación, debido a que la primera observación no tiene antecedentes. Para evitar dicha pérdida, la primera observación se transforma de la siguiente manera:[35] $Y_1\sqrt{1-\rho^2}$ y $X_1\sqrt{1-\rho^2}$. Esta transformación se conoce como **transformación Prais-Winsten**.

## Cuando no se conoce $\rho$

Aunque es sencillo aplicar la regresión en diferencias generalizada dada en (12.9.5), por lo general es difícil efectuarla en la práctica porque pocas veces se conoce $\rho$. Por consiguiente, se requieren formas de calcular $\rho$. Hay varias posibilidades.

### *Método de primeras diferencias*

Como $\rho$ se encuentra entre 0 y $\pm 1$, se puede partir de dos posiciones extremas. En un extremo, se puede suponer que $\rho = 0$, es decir, no hay correlación serial (de primer orden) y en el otro extremo, se puede considerar que $\rho = \pm 1$, es decir, autocorrelación positiva o negativa perfecta. En realidad, cuando se efectúa una regresión, suele suponerse que no hay autocorrelación y luego se deja que la prueba de Durbin-Watson u otras pruebas demuestren si el supuesto es justificado. Sin embargo, si $\rho = +1$, la ecuación en diferencias generalizada (12.9.5) se reduce a la **ecuación en primeras diferencias**, pues

$$Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$$

o bien

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \tag{12.9.7}$$

donde $\Delta$ es el operador de primeras diferencias introducido en (12.1.10).

Como el término de error en (12.9.7) está libre de la correlación serial (de primer orden) (¿por qué?), todo lo que hay que hacer para llevar a cabo la regresión (12.9.7) es formar las primeras diferencias de la regresada y la(s) regresora(s), y realizar la regresión sobre esas primeras diferencias.

La transformación de primeras diferencias puede resultar adecuada si el coeficiente de autocorrelación es muy alto, por ejemplo, superior a 0.8; o si el $d$ de Durbin-Watson es muy bajo. Maddala propuso esta rudimentaria regla práctica: *utilice la forma de primeras diferencias siempre que* $d < R^2$.[36] Esto es lo que sucede en la regresión de los salarios sobre la productividad (12.5.1), en donde se tiene que $d = 0.2176$ y $r^2 = 0.9845$. En breve presentaremos la regresión de primeras diferencias para el ejemplo ilustrativo.

Una característica importante del modelo de primeras diferencias (12.9.7) es que **carece de intercepto**. Por tanto, para estimar (12.9.7) se tiene que emplear la rutina de la **regresión a través del origen** (es decir, suprimir el término del intercepto), la cual ahora está disponible en la mayoría de los software. Sin embargo, si olvida eliminar el término del intercepto en el modelo y estima el siguiente modelo que incluye dicho término

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t \tag{12.9.8}$$

[35] La pérdida de una observación quizá no sea algo preocupante en muestras grandes, pero puede significar una diferencia fundamental en los resultados de muestras pequeñas. Si no se transforma la primera observación como se indica, la varianza de error no será homoscedástica. Para saber más sobre lo anterior, véase Jeffrey Wooldridge, *op. cit.*, p. 388. Para conocer algunos resultados Monte Carlo sobre la importancia de la primera observación, véase Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, tabla 10.1, p. 349.

[36] Maddala, *op. cit.*, p. 232.

entonces el modelo original debe tener una *tendencia* y $\beta_1$ representa el coeficiente de la variable de tendencia.[37] En consecuencia, un beneficio "accidental" de introducir el término del intercepto en el modelo de primeras diferencias es probar la presencia de una variable de tendencia en el modelo original.

En cuanto a la regresión de salarios sobre productividad (12.5.2), con el esquema AR(1) y un valor $d$ bajo en relación con $r^2$, volvemos a hacer la regresión (12.5.2) en la forma de primeras diferencias, sin el término de intercepto; recuerde que (12.5.2) está en *forma de nivel*. Los resultados son los siguientes:[38]

$$\begin{aligned} \widehat{\Delta Y}_t &= 0.6539 \Delta X_t \\ t &= (11.4042) \qquad r^2 = 0.4264 \qquad d = 1.7442 \end{aligned} \tag{12.9.9}$$

En comparación con la regresión de forma de nivel (12.5.2), observamos que el coeficiente de pendiente no cambió mucho, pero el valor $r^2$ disminuyó de manera considerable. Esto sucede porque al tomar las primeras diferencias se estudia en esencia el comportamiento de variables alrededor de sus valores de tendencia (lineal). Por supuesto, no podemos comparar la $r^2$ en forma directa de (12.9.9) con la de (12.5.2) porque las variables dependientes en los dos modelos son distintas.[39] Asimismo, observe que, en comparación con la regresión original, el valor $d$ aumentó de manera drástica, con lo que quizá se señale que hay muy poca autocorrelación en la regresión de primeras diferencias.[40]

Otro aspecto interesante de la transformación de primeras diferencias se relaciona con las propiedades estacionarias de las series de tiempo subyacentes. Considere de nuevo la ecuación (12.2.1), que describe el esquema AR(1). Ahora bien, si de hecho $\rho = 1$, resulta claro de las ecuaciones (12.2.3) y (12.2.4) que la serie $u_t$ es *no estacionaria*, pues las varianzas y covarianzas se vuelven infinitas. Es por esta razón que, cuando estudiamos este tema, pusimos la restricción de que $|\rho| < 1$. Pero es evidente, a partir de (12.2.1), que si el coeficiente de autocorrelación es de hecho 1, entonces (12.2.1) se convierte en

$$u_t = u_{t-1} + \varepsilon_t$$

o

$$(u_t - u_{t-1}) = \Delta u_t = \varepsilon_t \tag{12.9.10}$$

Es decir, es $u_t$ en su forma de primeras diferencias la que se vuelve estacionaria, porque es igual a $\varepsilon_t$, un término de error de ruido blanco.

Lo que se debe destacar del análisis anterior es que si las series de tiempo originales son no estacionarias, a menudo sus primeras diferencias se convierten en estacionarias. Por consiguiente, la transformación de primeras diferencias tiene un doble propósito: puede eliminar la autocorrelación (de primer orden) y también causar que una serie de tiempo sea estacionaria. En la **parte 5** analizaremos ese tema, y también la econometría del análisis de series de tiempo, con cierto detalle.

Ya mencionamos que la transformación de primeras diferencias resulta adecuada si $\rho$ es alta o si $d$ es baja. En estricto sentido, la trasformación de primeras diferencias es válida sólo si $\rho = 1$.

[37] Es fácil demostrarlo. Sea $Y_t = \alpha_1 + \beta_1 t + \beta_2 X_t + u_t$. Por tanto, $Y_{t-1} = \alpha + \beta_1(t-1) + \beta_2 X_{t-1} + u_{t-1}$. Al restar la última de la primera, obtendrá: $\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t$, lo cual muestra que el término del intercepto en la ecuación es en realidad el coeficiente de la variable de tendencia en el modelo original. Recuerde que suponemos que $\rho = 1$.

[38] En el ejercicio 12.38 se le pide hacer la regresión de este modelo, incluyendo su término constante.

[39] La comparación de $r^2$ en la forma de nivel y de primeras diferencias es un poco complicada. Para un análisis amplio de lo anterior, véase Maddala, *op. cit.*, capítulo 6.

[40] No resulta claro si el valor $d$ calculado en la regresión de primeras diferencias se puede interpretar del mismo modo que el original (la regresión en su forma de nivel). Sin embargo, al aplicar la prueba de las rachas, se observa si no hay evidencia de autocorrelación en los residuos de la regresión de primeras diferencias.

De hecho, existe una prueba, llamada **prueba de Berenblutt-Webb**,[41] para probar la hipótesis de que $\rho = 1$. El estadístico de prueba se llama **estadístico g**, y se define como sigue:

$$g = \frac{\sum_2^n \hat{e}_t^2}{\sum_1^n \hat{u}_t^2} \qquad \textbf{(12.9.11)}$$

donde $\hat{u}_t$ son los residuos de MCO de la regresión original (es decir, en su forma de nivel) y $\varepsilon_t$ son los residuos de MCO de la regresión de primeras diferencias. Tenga en cuenta que en la forma de primeras diferencias no hay intercepto.

Para probar la significancia del estadístico $g$, suponiendo que la regresión de la forma de nivel contiene el término de intercepto, se utilizan las tablas Durbin-Watson, salvo que la hipótesis nula ahora es que $\rho = 1$ en vez de la hipótesis de Durbin-Watson de que $\rho = 0$.

Retome la regresión de los salarios sobre la productividad, y para la regresión original (12.5.2) obtenemos $\sum \hat{u}_t^2 = 0.0214$ y $\sum \hat{e}_t^2 = 0.0046$. Al sustituir estos valores en el estadístico $g$ dado en (12.9.11), obtenemos

$$g = \frac{0.0046}{0.0214} = 0.2149 \qquad \textbf{(12.9.12)}$$

Al consultar la tabla de Durbin-Watson para 45 observaciones (el número más cercano a 45 observaciones) y 1 variable explicativa (apéndice D, tabla D.5), vemos que $d_L = 1.288$ y $d_U = 1.376$ (en el nivel de significancia de 5%). Como el valor $g$ observado se encuentra por debajo del límite inferior de $d$, no rechazamos la hipótesis que el verdadero $\rho = 1$. *Tenga presente que, aunque se utilizan las mismas tablas de Durbin-Watson, ahora la hipótesis nula es que $\rho = 1$ y no que $\rho = 0$.* En vista de este hallazgo, los resultados dados en (12.9.9) pueden ser aceptables.

### *$\rho$ basada en el estadístico d de Durbin-Watson*

Si no podemos utilizar la transformación de primeras diferencias porque $\rho$ no está lo bastante cerca de la unidad, hay un método fácil para estimarla a partir de la relación establecida con anterioridad entre $d$ y $\rho$ en (12.6.10), de la cual $\rho$ se estima de la siguiente forma:

$$\hat{\rho} \approx 1 - \frac{d}{2} \qquad \textbf{(12.9.13)}$$

Por tanto, en muestras razonablemente grandes, se puede obtener $\rho$ de (12.9.13) y emplearla para transformar los datos, como vimos en la ecuación en diferencias generalizadas (12.9.5). Tenga presente que la relación entre $\rho$ y $d$ dada en (12.9.13) quizá no sea válida para muestras pequeñas; por eso, Theil y Nagar propusieron una modificación, que se presenta en el ejercicio 12.6.

En la regresión de los salarios sobre la productividad (12.5.2) obtuvimos un valor $d$ igual a 0.2176; con dicho valor en (12.9.13) obtuvimos $\hat{\rho} \approx 0.8912$. Mediante este valor estimado de $\rho$ calculamos la regresión (12.9.5). Todo lo que se tiene que hacer es restar la cantidad 0.8912, multiplicada por el valor anterior de $Y$, de su valor actual, y de manera similar, restar 0.8912, multiplicada por el valor anterior de $X$, de su valor actual y realizar la regresión MCO sobre las variables así transformadas, como en (12.9.6), donde $Y_t^* = (Y_t - 0.8912Y_{t-1})$ y $X_t^* = (X_t - 0.8912X_{t-1})$.

### *$\rho$ estimada a partir de los residuos*

Si el esquema AR(1) $u_t = \rho u_{t-1} + \varepsilon_t$ es válido, una manera sencilla de estimar $\rho$ es hacer la regresión de los residuos $\hat{u}_t$ o $\hat{u}_{t-1}$ de modo que las $\hat{u}_t$ sean estimadores consistentes de las verdaderas $u_t$, como ya vimos. Es decir, hacemos la siguiente regresión:

$$\hat{u}_t = \rho \, . \, \hat{u}_{t-1} + v_t \qquad \textbf{(12.9.14)}$$

[41] I.I. Berenblutt y G.I. Webb, "A New Test for Autocorrelated Errors in the Linear Regression Model", *Journal of the Royal Statistical Society*, Serie B, vol. 35, núm. 1, 1973, pp. 33-50.

donde $\hat{u}_t$, son los residuos obtenidos de la regresión original (en su forma de nivel) y $v_t$ son el término de error de esta regresión. Observe que no hay necesidad de introducir el término de intercepto en (12.9.14), porque sabemos que los residuos de MCO suman cero.

Los residuos de la regresión de los salarios sobre la producción dados en (12.5.1) ya se calcularon en la tabla 12.5. Con esos resultados obtuvimos lo siguiente:

$$\begin{aligned} \widehat{\hat{u}_t} &= \quad 0.8678\hat{u}_{t-1} \\ t &= (12.7359) \qquad r^2 = 0.7863 \end{aligned} \tag{12.9.15}$$

Como muestra la regresión, $\hat{\rho} = 0.8678$. Con esta estimación transformamos el modelo original a (12.9.6). Como la $\rho$ estimada mediante este procedimiento es casi la misma que la obtenida por el $d$ de Durbin-Watson, los resultados de la regresión al utilizar la $\rho$ de (12.9.15) no deben ser muy distintos de los obtenidos mediante la $\rho$ estimada con el $d$ de Durbin-Watson. Dejamos al lector comprobar lo anterior.

### *Métodos iterativos para estimar $\rho$*

Todos los métodos para estimar $\rho$ que hemos visto proporcionan sólo una estimación de $\rho$. Pero existen los llamados **métodos iterativos** que estiman $\rho$ de manera iterativa, es decir, mediante aproximaciones sucesivas, comenzando con algún valor inicial de $\rho$. Entre estos métodos, mencionaremos los siguientes: **procedimiento iterativo de Cochrane-Orcutt, procedimiento de dos pasos de Cochrane-Orcutt**, **procedimiento de dos pasos de Durbin** y **procedimiento de rastreo o de búsqueda de Hildreth-Lu**. El más popular es el método iterativo de Cochrane-Orcutt. Para ahorrar espacio, analizaremos los métodos iterativos mediante ejercicios. Recuerde que el último objetivo de tales métodos es proporcionar una estimación de $\rho$ con que se obtengan las estimaciones de los parámetros de MCG. Una ventaja del método iterativo de Cochrane-Orcutt es que sirve para estimar no sólo un esquema AR(1), sino también esquemas autorregresivos de orden superior, como $\hat{u}_t = \hat{\rho}_1\hat{u}_{t-1} + \hat{\rho}_2\hat{u}_{t-2} + v_t$, que son AR(2). Tras obtener dos $\rho$, se extiende con facilidad la ecuación en diferencias generalizada (12.9.6). Por supuesto, la computadora puede ahora efectuar lo anterior.

De regreso al ejemplo de la regresión de los salarios sobre la productividad y suponiendo un esquema AR(1), utilizamos el método iterativo Cochrane-Orcutt, el cual da las siguientes estimaciones de $\rho$: 0.8876, 0.9944 y 0.8827. El último valor, 0.8827, ahora sirve para transformar el modelo original como en (12.9.6) y se estima mediante MCO. Por supuesto, los MCO en el modelo transformado son simplemente los MCG. Los resultados son los siguientes:

El software Stata puede estimar los coeficientes del modelo, además de $\rho$. Por ejemplo, si suponemos AR(1), Stata produce los siguientes resultados:

$$\begin{aligned} \hat{Y}_t^* &= \ 43.1042 \ + \ 0.5712X_t \\ \text{ee} &= \ (4.3722) \quad (0.0415) \\ t &= \ (9.8586) \ \ (13.7638) \qquad r^2 = 0.8146 \end{aligned} \tag{12.9.16}$$

A partir de estos resultados, observamos que la rho estimada ($\hat{\rho}$) es $\approx 0.8827$, lo cual no difiere mucho de la $\hat{\rho}$ de la ecuación (12.9.15).

Como ya señalamos, en la ecuación en diferencia generalizada (12.9.6) se pierde una observación porque la primera no tiene antecedente. Para no perder la primera observación se usa la *transformación de Prais-Winsten.* Con esta transformación y STATA (versión 10) obtenemos los siguientes resultados de la regresión de salarios-productividad.

$$\begin{aligned} \text{Rcompb}_t &= \ 32.0434 \ + 0.6628 \ \text{Prodb}_t \\ \text{ee} &= \ (3.7182) \quad (0.0386) \qquad r^2 = 0.8799 \end{aligned} \tag{12.9.17}$$

En esta transformación, el valor $\rho$ fue 0.9193, que se obtuvo después de 13 iteraciones. *Debe señalarse que si la primera observación no se transforma a la manera de Prais-Winsten y se elimina, los resultados son a veces muy diferentes, en especial en muestras pequeñas.* Observe que el valor $\rho$ obtenido aquí no difiere mucho del que se obtuvo con la ecuación (12.9.15).

### *Comentarios generales*

Hay varios puntos que se deben destacar respecto de la corrección de la autocorrelación mediante los diversos métodos analizados.

*En primer lugar*, como para muestras grandes los estimadores de MCO son consistentes, sin importar la autocorrelación, no afecta en nada si se estima $\rho$ a partir del estadístico $d$ de Durbin-Watson, de la regresión de los residuos del periodo actual sobre los residuos del periodo anterior o del procedimiento iterativo Cochrane-Orcutt, pues todos proporcionan estimados consistentes con la verdadera $\rho$. *En segundo lugar*, los distintos métodos analizados son básicamente métodos de dos pasos. En el primer paso se obtiene una estimación de la $\rho$ desconocida, y en el segundo se utiliza dicha estimación para transformar las variables a fin de calcular la ecuación en diferencias generalizada, que es básicamente MCG. Pero en vista de que se utiliza $\hat{\rho}$ en vez de la verdadera $\rho$, todos estos métodos de estimación se conocen en la bibliografía como métodos **MCG factibles (MCGF**) o **MCG estimados (MCGE**).

*En tercer lugar*, es importante observar que siempre que se estimen los parámetros del modelo transformado con un método **MCGF** o un **MCGE**, los coeficientes estimados no necesariamente tendrán las propiedades óptimas usuales del modelo clásico, como ser MELI, sobre todo en muestras pequeñas. Sin adentrarnos en complejidades técnicas, podemos enunciar, *como principio general, que siempre que se utilice un estimador en lugar de su verdadero valor, los coeficientes de MCO estimados quizá presenten las propiedades óptimas usuales en forma asintótica; es decir, para muestras grandes. Asimismo, los procedimientos convencionales para pruebas de hipótesis son, en estricto sentido, válidos de modo asintótico. En consecuencia, para muestras pequeñas, se debe tener cuidado al interpretar los resultados estimados.*

*En cuarto lugar*, al utilizar MCGF, si no se incluye la primera observación (como se hizo al principio con el procedimiento Cochrane-Orcutt), se pueden ver afectados de modo adverso no sólo los valores numéricos, sino también la eficiencia de los estimadores, sobre todo si el tamaño de la muestra es pequeño y las regresoras no son, estrictamente hablando, no estocásticas.[42] Por tanto, en pequeñas muestras es importante conservar la primera observación *al estilo* Prais-Winsten. Desde luego, si el tamaño de la muestra es razonablemente grande, el MCGF, con o sin primera observación, proporciona resultados similares. A propósito, en la bibliografía, el MCGF con transformación Prais-Winsten se conoce como **MCGF completo** o **MCGFC**.

## 12.10 El método Newey-West para corregir los errores estándar de MCO

En lugar de los métodos MCGF analizados en la sección anterior, podemos conservar los MCO pero con los errores estándar corregidos por autocorrelación, mediante un procedimiento desarrollado por Newey y West.[43] Se trata de una generalización de los errores estándar consistentes con heteroscedasticidad de White, los cuales examinamos en el capítulo anterior. Los errores estándar corregidos se conocen como **errores estándar CHA (consistentes con heteroscedasticidad y autocorrelación)**, o simplemente **errores Newey-West**. No abordaremos las matemáticas

[42] Esto es especialmente cierto si las regresoras muestran una tendencia, lo cual es muy común en los datos económicos.

[43] W.K. Newey y K. West, "A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, vol. 55, 1987, pp. 703-708.

propias del procedimiento Newey-West, pues son complejas,[44] pero la mayoría de los programas estadísticos modernos calcula los errores estándar Newey-West. Sin embargo, resulta importante destacar que el procedimiento Newey-West es, *en estricto sentido, válido para muestras grandes*, aunque quizá no sea adecuado para muestras pequeñas. No obstante, para muestras grandes ahora hay un método que origina errores estándar corregidos por autocorrelación, así que no hay que preocuparse por las transformaciones MCGF analizadas en capítulos anteriores. Por tanto, si una muestra es razonablemente grande, debe utilizarse el procedimiento Newey-West para corregir los errores estándar de MCO, no sólo para situaciones de autocorrelación, sino también para casos de heteroscedasticidad, pues el método CHA puede abordar ambos casos, a diferencia del método White, diseñado específicamente para la heteroscedasticidad.

Veamos una vez más el ejemplo de la regresión de salarios sobre productividad (12.5.1). Sabemos que esta regresión muestra autocorrelación. La muestra de 46 observaciones es razonablemente grande, por lo que podemos utilizar el procedimiento CHA. Con *EViews 4* obtenemos los siguientes resultados de la regresión:

$$
\begin{aligned}
\hat{Y}_t &= 32.7419 \;+\; 0.6704X_t \\
\text{ee} &= (2.9162)^* \quad (0.0302)^* \qquad\qquad (12.10.1)\\
&\qquad\qquad r^2 = 0.9765 \qquad d = 0.1719
\end{aligned}
$$

donde * denota errores estándar CHA.

Al comparar esta regresión con (12.5.1) encontramos que en ambas ecuaciones los coeficientes estimados y el valor $r^2$ son iguales. Pero es importante observar que los errores estándar CHA son mucho mayores que los errores estándar de MCO, y por tanto las razones *t* CHA son mucho más pequeñas que las razones *t* de MCO. Lo anterior demuestra que en realidad MCO subestimó los errores estándar verdaderos. Resulta curioso que el estadístico *d* para los modelos (12.5.1) y (12.10.1) sea el mismo. Pero no hay de qué preocuparse, pues el procedimiento CHA ya tomó esto en cuenta al corregir los errores estándar MCO.

## 12.11 MCO *versus* MCGF y CHA

El problema práctico que enfrenta el investigador es: en presencia de autocorrelación, los estimadores de MCO —aunque no están sesgados, son consistentes y están distribuidos de manera normal y asintótica—, no son eficientes. En consecuencia, ya no es adecuado el procedimiento de inferencia usual basado en las pruebas *t*, *F* y $\chi^2$. Por otra parte, los procedimientos MCGF y CHA producen estimadores eficientes, pero sus propiedades finitas, o para muestras pequeñas, no están bien documentadas. Esto significa que, para muestras pequeñas, MCGF y CHA en realidad pueden resultar peores que los MCO. De hecho, en un estudio Monte Carlo, Griliches y Rao[45] encontraron que si la muestra es relativamente pequeña y el coeficiente de autocorrelación $\rho$ es menor que 0.3, MCO es igual o mejor que MCGF. Así, como cuestión práctica, se puede utilizar MCO para muestras pequeñas en las que el $\rho$ estimado sea, por ejemplo, menor que 0.3. Por supuesto, siempre es relativo afirmar qué es pequeño y qué es grande, así que aquí entra el sentido común. Si sólo hay 15 o 20 observaciones, la muestra puede ser pequeña, pero si en cambio existen 50 o más observaciones, la muestra puede ser razonablemente grande.

[44] Si el lector conoce el álgebra matricial, el método se analiza en la obra de Greene, *op. cit.*, 4a. ed., pp. 462-463.

[45] Z. Griliches y P. Rao, "Small Simple Properties of Several Two-stage Regression Methods in the Context of Autocorrelated Errors", *Journal of the American Statistical Association*, vol. 64, 1969, pp. 253-272.

## 12.12 Otros aspectos de la autocorrelación

### Variables dicótomas y autocorrelación

En el capítulo 9 analizamos los modelos de regresión con una variable dicótoma. En particular, recuerde el modelo de regresión del ahorro sobre el ingreso estadounidenses de 1970 a 1995, que presentamos en (9.5.1), y que reproducimos a continuación:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \qquad \textbf{(12.12.1)}$$

donde $Y$ = ahorro
$X$ = ingreso
$D = 1$ para observaciones de 1982 a 1995
$D = 0$ para observaciones de 1970 a 1981

Los resultados de la regresión basados en este modelo se dan en (9.5.4). Por supuesto, este modelo se estimó con los supuestos usuales para MCO.

Pero ahora suponga que $u_t$ sigue un esquema autorregresivo de primer orden, AR(1). Es decir, $u_t = \rho u_{t-1} + \varepsilon_t$. Por lo común, si se conoce $\rho$ o se calcula mediante uno de los métodos vistos antes, podemos utilizar el método de diferencias generalizado para calcular los parámetros del modelo que no presenta autocorrelación (de primer orden). No obstante, la presencia de la variable dicótoma $D$ plantea un problema: observe que la variable dicótoma simplemente clasifica una observación como perteneciente al primero o al segundo periodo. ¿Cómo la transformamos? Seguimos el siguiente procedimiento:[46]

1. En (12.12.1), los valores de $D$ son cero para todas las observaciones del primer periodo; en el periodo 2, el valor de $D$ para la **primera** observación es $1/(1-\rho)$ en vez de 1, y es 1 para las demás observaciones.
2. La variable $X_t$ se transforma como $(X_t - \rho X_{t-1})$. Observe que se pierde una observación en esta transformación, a menos que se recurra a la **transformación de Prais-Winsten** para la primera observación, como se observó antes.
3. El valor de $D_t X_t$ es cero para todas las observaciones en el primer periodo (*nota:* $D_t$ es cero en el primer periodo); en el segundo periodo, la primera observación toma el valor de $D_t X_t = X_t$ y las observaciones restantes en el segundo periodo se igualan a $(D_t X_t - D_t \rho X_{t-1}) = (X_t - \rho X_{t-1})$. (*Nota:* el valor de $D_t$ en el segundo periodo es 1.)

Como señala el análisis anterior, la *observación crítica* es la primera observación del segundo periodo. Si se tiene cuidado en la forma que aquí se acaba de indicar, no debe haber problema al estimar regresiones como (12.12.1), sujetas a la autocorrelación AR(1). En el ejercicio 12.37 se pide al lector llevar a cabo esa transformación para los datos de ahorro e ingreso en Estados Unidos del capítulo 9.

### Modelos ARCH y GARCH

Justo como el término de error $u$ en el tiempo $t$ puede estar correlacionado con el término de error en el tiempo $(t-1)$ en un esquema AR(1), o con varios términos de error rezagados en el esquema general AR($p$), ¿puede existir autocorrelación en la varianza $\sigma^2$ en el tiempo $t$, respecto de sus valores rezagados uno o más periodos? Los investigadores que trabajan en el pronóstico de series de tiempo financieras, como precios accionarios, tasas de inflación y tasas de cambio de divisas, han observado dicha autocorrelación. A tal autocorrelación se le han dado nombres más bien intimidantes, como **heteroscedasticidad condicional autorregresiva (ARCH)**, si la varianza del error está relacionada con el término del error al cuadrado en el periodo anterior, y **heteroscedasticidad condicional autorregresiva generalizada (GARCH)**, si la varianza del

[46] Véase Maddala, *op. cit.*, pp. 321-322.

error está relacionada con los términos del error al cuadrado de varios periodos en el pasado. Como este tema pertenece al área general de la econometría de las series de tiempo, lo veremos con mayor profundidad en los capítulos correspondientes. El objetivo aquí es señalar que la autocorrelación no sólo abarca las relaciones entre los términos del error actuales y anteriores, sino también las varianzas de los errores actuales y anteriores.

### Coexistencia de la autocorrelación y la heteroscedasticidad

¿Qué pasa si un modelo de regresión presenta heteroscedasticidad y autocorrelación? ¿Podemos resolver este problema de manera secuencial, es decir, abordar primero la heteroscedasticidad y luego la autocorrelación? De hecho, un autor afirma que: "La autocorrelación sólo se puede detectar después de controlar la heteroscedastiacidad".[47] Pero, ¿podemos desarrollar una prueba omnipotente que resuelva de manera simultánea esos problemas y otros más (por ejemplo, la especificación de modelo)? Sí, tales pruebas existen, pero su estudio trasciende los objetivos de este libro, por lo que lo dejamos para consulta en las referencias.[48] Sin embargo, como ya señalamos, podemos usar los errores estándar CHA, pues toman en cuenta tanto la autocorrelación como la heteroscedasticidad, siempre que la muestra sea razonablemente grande.

## 12.13 Ejemplo para concluir

En el ejemplo 10.2 presentamos datos sobre consumo, ingreso, riqueza y tasas de interés en Estados Unidos, todos en términos reales. Con base en estos datos estimamos la siguiente función de consumo para Estados Unidos de 1947 a 2000, con la regresión del logaritmo de consumo sobre los logaritmos de ingreso y de riqueza. No expresamos la tasa de interés en forma de logaritmo porque algunas cifras relativas a la tasa de interés real eran negativas.

Variable dependiente: ln(CONSUMO)
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | -0.467711 | 0.042778 | -10.93343 | 0.0000 |
| ln(INGRESO) | 0.804873 | 0.017498 | 45.99836 | 0.0000 |
| ln(RIQUEZA) | 0.201270 | 0.017593 | 11.44060 | 0.0000 |
| INTERÉS | -0.002689 | 0.000762 | -3.529265 | 0.0009 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.999560 | Media de la variable dependiente | 7.826093 |
| R cuadrada ajustada | 0.999533 | Desv. est. de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.011934 | Estadístico *F* | 37832.59 |
| Suma de cuadrados residual | 0.007121 | Probabilidad (estadístico *F*) | 0.000000 |
| Log verosimilitud | 164.5880 | Estadístico de Durbin-Watson | 1.289219 |

Como era de esperar, las elasticidades del ingreso y la riqueza son positivas, y la semielasticidad de la tasa de interés, negativa. Aunque al parecer los coeficientes estimados son muy significativos estadísticamente en lo individual, se precisa una inspección para detectar una posible autocorrelación en el término de error. Como sabemos, en presencia de autocorrelación, pueden

[47] Lois W. Sayrs, *Pooled Time Series Analysis*, Sage, California, 1989, p. 19.

[48] Véase Jeffrey M. Wooldridge, *op. cit.*, pp. 402-403, y A.K. Bera y C.M. Jarque, "Efficient Tests for Normalita, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence", *Economic Letters*, vol. 7, 1981, pp. 313-318.

subestimarse los errores estándar estimados. Al examinar el estadístico *d* de Durbin-Watson, parece que en los términos de error de la función de consumo hay autocorrelación (de primer grado) (verifíquelo).

Para confirmar esto, estimamos la función de consumo teniendo en cuenta la autocorrelación AR(1). Los resultados son los siguientes:

```
Variable dependiente: lnCONSUMO
Método: Mínimos cuadrados
Muestra (ajustada): 1948-2000
Observaciones incluidas: 53 después de los ajustes
Convergencia lograda después de 11 iteraciones
```

| | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| C | -0.399833 | 0.070954 | -5.635112 | 0.0000 |
| lnINGRESO | 0.845854 | 0.029275 | 28.89313 | 0.0000 |
| lnRIQUEZA | 0.159131 | 0.027462 | 5.794501 | 0.0000 |
| INTERÉS | 0.001214 | 0.000925 | 1.312986 | 0.1954 |
| AR(1) | 0.612443 | 0.100591 | 6.088462 | 0.0000 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.999688 | Media de la variable dependiente | 7.843871 |
| R cuadrada ajustada | 0.999662 | Desviación estándar de la variable dependiente | 0.541833 |
| Error estándar de la regresión | 0.009954 | Estadístico $F$ | 38503.91 |
| Suma de cuadrados residual | 0.004756 | Probabilidad (estadístico $F$) | 0.000000 |
| Log verosimilitud | 171.7381 | Estadístico de Durbin-Watson | 1.874724 |

Estos resultados muestran claramente la presencia de autocorrelación en la regresión. Se deja al lector la tarea de eliminar la autocorrelación mediante alguna de las transformaciones analizadas en este capítulo. Puede usar la $\rho$ estimada de 0.6124 para las transformaciones. A continuación presentamos los resultados basados en errores estándar Newey-West (CHA) que toman en cuenta la autocorrelación.

```
Variable dependiente: lnCONSUMO
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54
Errores estándar Newey-West CHA y covarianza (truncamiento de rezago = 3)
```

| | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| C | -0.467714 | 0.043937 | -10.64516 | 0.0000 |
| INGRESO | 0.804871 | 0.017117 | 47.02132 | 0.0000 |
| RIQUEZA | 0.201272 | 0.015447 | 13.02988 | 0.0000 |
| INTERÉS | 0.002689 | 0.000880 | -3.056306 | 0.0036 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.999560 | Media de la variable dependiente | 7.826093 |
| R cuadrada ajustada | 0.999533 | Desviación estándar de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.011934 | Estadístico $F$ | 37832.71 |
| Suma de cuadrados residual | 0.007121 | Probabilidad (estadístico $F$) | 0.000000 |
| | | Estadístico de Durbin-Watson | 1.289237 |

La diferencia principal entre la primera y la última de las regresiones anteriores es que los errores estándar de los coeficientes estimados cambiaron de manera considerable. Pese a ello, los coeficientes estimados de las pendientes son aún muy significativos estadísticamente. Sin embargo, no hay garantía de que siempre será así.

## Resumen y conclusiones

1. Si se viola el supuesto del modelo clásico de regresión lineal de que los errores o las perturbaciones $u_t$ consideradas dentro de la función de regresión poblacional (FRP) son aleatorios o no correlacionados, surge el problema de autocorrelación o correlación serial.
2. La autocorrelación surge por diversas razones, como la inercia o pasividad de las series de tiempo económicas, el sesgo de especificación resultante de excluir variables importantes del modelo o de utilizar la forma funcional incorrecta, el fenómeno de la telaraña, el manejo y transformación de datos, etc. Como resultado, es útil distinguir entre la autocorrelación pura y la autocorrelación "inducida", debido a uno o más de los factores que acabamos de mencionar.
3. Aunque en presencia de autocorrelación los estimadores de MCO se mantienen insesgados, consistentes y distribuidos asintóticamente en forma normal, dejan de ser eficientes. Como resultado, las pruebas $\chi^2$, $t$ y $F$ usuales no son aplicables legítimamente. Por tanto, se necesita la aplicación de medidas correctivas.
4. El remedio depende de la naturaleza de la interdependencia entre las perturbaciones $u_t$. Pero como las $u_t$ no son observables, la práctica común es suponer que algún mecanismo las generó.
5. El mecanismo más común es el esquema autorregresivo de primer orden de Markov, que supone que la perturbación en el tiempo actual está linealmente relacionada con el término de perturbación en el tiempo anterior, el coeficiente de autocorrelación $\rho$ que da el grado de interdependencia. Este mecanismo se conoce como esquema AR(1).
6. Si el esquema AR(1) es válido y se conoce el coeficiente de autocorrelación, el problema de correlación serial se resuelve fácilmente mediante la transformación de los datos según el procedimiento de diferencias generalizado. El esquema AR(1) se generaliza sin dificultad a un esquema AR(p). También se puede suponer un mecanismo de promedios móviles (PM) o una mezcla de los esquemas AR y PM, conocido como ARMA. Estudiaremos este tema en los capítulos sobre econometría de series de tiempo.
7. Aunque utilicemos un esquema AR(1), el coeficiente de autocorrelación $\rho$ no se conoce *a priori*. Consideramos diversos métodos para estimar $\rho$, como el $d$ de Durbin-Watson, el $d$ modificado de Theil-Nagar, el procedimiento de dos etapas de Cochrane-Orcutt (C-O), el procedimiento iterativo C-O y el método de dos etapas de Durbin. En muestras grandes, estos métodos suelen producir estimaciones similares de $\rho$, aunque en muestras pequeñas tienen un desempeño diferente. En la práctica, el método iterativo C-O ha cobrado gran popularidad.
8. Con cualquiera de los métodos que acabamos de estudiar, podemos utilizar el método de diferencias generalizado para calcular los parámetros del modelo transformado mediante MCO, que en esencia es lo mismo que MCG. Pero en vista de que se estima $\rho$ $(=\hat{\rho})$, este método de estimación se conoce como factible, o estimado, y se abrevia MCG, MCGF o MCGE.
9. Al utilizar MCGE, se debe tener cuidado al eliminar la primera observación, pues en muestras pequeñas la inclusión o exclusión de la primera observación puede influir de manera drástica en los resultados. Por tanto, para muestras pequeñas, es aconsejable transformar la primera observación de acuerdo con el procedimiento Prais-Winsten. Sin embargo, en muestras grandes no importa si se incluye o excluye la primera observación.
10. Es muy importante notar que el método de MCGE presenta las propiedades estadísticas óptimas usuales sólo en muestras grandes. Para muestras pequeñas, el método de MCO puede resultar realmente mejor que el MCGE, sobre todo si $\rho < 0.3$.
11. En lugar de utilizar MCGE, todavía se pueden usar MCO, pero corrigiendo la autocorrelación de los errores estándar mediante el procedimiento CHA de Newey-West. En estricto sentido, este procedimiento es válido para muestras grandes. Una ventaja es que no sólo corrige la autocorrelación, sino también la heteroscedasticidad, en su caso.

12. Por supuesto, antes del remedio está la detección de la autocorrelación. Existen métodos formales e informales de detección. Entre los informales está el de simplemente graficar los residuos estandarizados o reales, o graficar los residuos reales respecto de los residuos anteriores. Entre los métodos formales se encuentran la prueba de rachas, la prueba *d* de Durbin-Watson, la de normalidad asintótica, la de Berenblutt-Webb y la de Breusch-Godfrey (BG). De todas, la más popular es la prueba *d* de Durbin-Watson. A pesar de su ilustre pasado, esta prueba tiene graves limitaciones. Es mejor la prueba BG, pues es más general debido a que permite las estructuras de error AR y PM, así como la presencia de la regresada rezagada como variable explicativa. Pero tenga en cuenta que es una prueba para muestras grandes.
13. En este capítulo también analizamos de manera muy sucinta la detección de la autocorrelación en presencia de variables dicótomas.

## EJERCICIOS

### Preguntas

12.1. Establezca si las siguientes afirmaciones son verdaderas o falsas. Justifique su respuesta brevemente.

*a*) Cuando hay presencia de autocorrelación, los estimadores de MCO son sesgados e ineficientes.

*b*) La prueba *d* de Durbin-Watson supone que la varianza del término de error $u_t$ es homoscedástica.

*c*) La transformación de primeras diferencias para eliminar la autocorrelación supone que el coeficiente de autocorrelación $\rho$ es $-1$.

*d*) Los valores $R^2$ de dos modelos, de los cuales uno corresponde a una regresión en forma de primeras diferencias y el otro a una regresión en su forma de nivel, no son directamente comparables.

*e*) Un *d* de Durbin-Watson significativo no necesariamente denota autocorrelación de primer orden.

*f*) En presencia de autocorrelación, las varianzas calculadas convencionalmente y los errores estándar de los valores pronosticados son ineficientes.

*g*) La exclusión de una o varias variables importantes de un modelo de regresión puede producir un valor *d* significativo.

*h*) En el esquema AR(1), una prueba de hipótesis de que $\rho = 1$ puede hacerse mediante el estadístico *g* de Berenblutt-Webb, lo mismo que con el estadístico *d* de Durbin-Watson.

*i*) En la regresión de primeras diferencias de *Y* sobre primeras diferencias de *X*, si hay un término constante y un término de tendencia lineal, significa que en el modelo original hay un término de tendencia lineal y uno de tendencia cuadrática.

12.2. Con una muestra de 50 observaciones y 4 variables explicativas, ¿qué puede decir sobre autocorrelación si *a*) $d = 1.05$, *b*) $d = 1.40$, *c*) $d = 2.50$ y *d*) $d = 3.97$?

12.3. Al estudiar el movimiento en la participación de la producción de los trabajadores en el valor agregado (es decir, la participación laboral), Gujarati* consideró los siguientes modelos:

$$\text{Modelo A:} \quad Y_t = \beta_0 + \beta_1 t + u_t$$

$$\text{Modelo B:} \quad Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$$

* Damodar Gujarati, "Labor's Share in Manufacturing Industries", *Industrial and Labor Relations Review*, vol. 23, núm. 1, octubre de 1969, pp. 65-75.

donde $Y$ = participación laboral y $t$ = tiempo. Con base en información anual de 1949 a 1964 se obtuvieron los siguientes resultados para la industria metalúrgica básica:

$$\text{Modelo A:}\quad \hat{Y}_t = 0.4529 - \underset{(-3.9608)}{0.0041t} \qquad R^2 = 0.5284 \qquad d = 0.8252$$

$$\text{Modelo B:}\quad \hat{Y}_t = 0.4786 - \underset{(-3.2724)}{0.0127t} + \underset{(2.7777)}{0.0005t^2}$$

$$R^2 = 0.6629 \qquad d = 1.82$$

donde las cifras entre paréntesis son las razones $t$.

*a*) ¿Hay correlación serial en el modelo A? ¿En el modelo B?

*b*) ¿Qué explica la correlación serial?

*c*) ¿Cómo distinguiría entre autocorrelación "pura" y sesgo de especificación?

12.4. *Detección de la autocorrelación: prueba de la razón de von Neumann*.* Suponiendo que los residuos $\hat{u}_t$ se obtienen aleatoriamente de una distribución normal, von Neumann demostró que para *n grande*, la razón

$$\frac{\delta^2}{s^2} = \frac{\sum(\hat{u}_i - \hat{u}_{i-1})^2/(n-1)}{\sum(\hat{u}_i - \bar{\hat{u}})^2/n} \qquad \textit{Nota}\text{: } \bar{\hat{u}} = 0 \text{ en MCO}$$

llamada **razón de von Neumann**, tiene una distribución aproximadamente normal con media

$$E\,\frac{\delta^2}{s^2} = \frac{2n}{n-1}$$

y varianza

$$\operatorname{var}\frac{\delta^2}{s^2} = 4n^2\,\frac{n-2}{(n+1)(n-1)^3}$$

*a*) Si $n$ es lo bastante grande, ¿cómo utilizaría la razón de von Neumann para probar la autocorrelación?

*b*) ¿Cuál es la relación entre la $d$ de Durbin-Watson y la razón de von Neumann?

*c*) El estadístico $d$ se encuentra entre 0 y 4. ¿Cuáles son los límites correspondientes para la razón de von Neumann?

*d*) Como la razón depende del supuesto de que las $\hat{u}$ se obtienen aleatoriamente de una distribución normal, ¿qué validez tiene este supuesto para los residuos de MCO?

*e*) Si en una aplicación se encontró que la razón era 2.88 con 100 observaciones; evalúe la hipótesis de que no hay correlación serial en los datos.

*Nota:* B.I. Hart tabuló los valores críticos de la razón de von Neumann para tamaños de muestras de hasta 60 observaciones.†

12.5. En una sucesión de 17 residuos, 11 positivos y 6 negativos, el número de rachas fue de 3. ¿Hay aquí evidencia de autocorrelación? ¿Cambiaría el resultado si hubiera 14 rachas?

* J. von Neumann, "Distribution of the Ratio of the Mean Square Successive Difference to the Variance", *Annals of Mathematical Statistics*, vol. 12, 1941, pp. 367-395.

† La tabla se encuentra en Johnston, *op. cit.*, 3a. ed., p. 559.

12.6. *Estimación de $\rho$ de Theil-Nagar basada en el estadístico d.* Theil y Nagar propusieron que, en muestras pequeñas, en lugar de estimar $\rho$ como $(1 - d/2)$ se estimara como

$$\hat{\rho} = \frac{n^2(1 - d/2) + k^2}{n^2 - k^2}$$

donde $n$ = número total de observaciones, $d = d$ de Durbin-Watson y $k$ = número de coeficientes que se van a estimar (incluso el intercepto).

Muestre que, para una $n$ grande, esta estimación de $\rho$ es igual a la obtenida por la fórmula más simple $(1 - d/2)$.

12.7. *Estimación de $\rho$: procedimiento de búsqueda o exploración de Hildreth-Lu.** Como en el esquema autorregresivo de primer orden

$$u_t = \rho u_{t-1} + \varepsilon_t$$

se espera que $\rho$ se encuentre entre $-1$ y $+1$, Hildreth y Lu proponen una "exploración" o procedimiento sistemático de búsqueda para localizarlo. Recomiendan seleccionar $\rho$ entre $-1$ y $+1$ con intervalos, por ejemplo, de 0.1 de unidad, y transformando los datos mediante la ecuación en diferencias generalizada (12.6.5). Así, se puede seleccionar $\rho$ de $-0.9, -0.8, \ldots, 0.8, 0.9$. Para cada $\rho$ seleccionada se efectúa la ecuación en diferencias generalizada y se obtiene la SCR asociada: $\sum \hat{u}_t^2$. Hildreth y Lu proponen seleccionar el valor de $\rho$ que minimice SCR (por tanto, se maximiza $R^2$). Si se requiere mayor refinación, se sugieren intervalos unitarios más pequeños, es decir, de 0.01 de unidad, como $-0.99, -0.98, \ldots, 0.90, 0.91$, y así sucesivamente.

*a*) ¿Cuáles son las ventajas del procedimiento Hildreth-Lu?

*b*) ¿Cómo se sabe que el valor $\rho$ seleccionado en última instancia para transformar los datos garantizará en realidad una $\sum \hat{u}_t^2$ mínima?

12.8. *Estimación de $\rho$: el procedimiento iterativo Cochrane-Orcutt (C-O).*† Como ilustración de este método, considere el modelo de dos variables:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (1)$$

y el esquema AR(1)

$$u_t = \rho u_{t-1} + \varepsilon_t, -1 < \rho < 1 \quad (2)$$

Cochrane y Orcutt recomendaron lo siguiente para estimar $\rho$.

1. Calcule (1) mediante la rutina usual de MCO y obtenga los residuos $\hat{u}_t$. A propósito, observe que puede haber más de una variable $X$ en el modelo.
2. Con los residuos calculados en el paso 1, haga la siguiente regresión:

$$\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + v_t \quad (3)$$

que es la contraparte empírica de (2).‡

3. Con $\hat{\rho}$ obtenida en (3), calcule la ecuación en diferencias generalizada (12.9.6).

* G. Hildreth y J.Y. Lu, "Demand Relations with Autocorrelated Disturbances", Michigan State University, *Agricultural Experiment Station*, Tech. Bull. 276, noviembre de 1960.

† D. Cochrane y G.H. Orcutt, "Applications of Least-Square Regressions to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association*, vol. 44, 1949, pp. 32-61.

‡ Observe que $\hat{\rho} = \sum \hat{u}_t \hat{u}_{t-1} / \sum \hat{u}_t^2$ (¿por qué?). Aunque sesgada, $\hat{\rho}$ es un estimador consistente de la verdadera $\rho$.

4. Como no se sabe *a priori* si la $\hat{\rho}$ obtenida de (3) es la mejor estimación de $\rho$, sustituya los valores de $\hat{\beta}_1^*$ y $\hat{\beta}_2^*$, del paso (3) para la regresión original (1), y obtenga los nuevos residuos, digamos, $\hat{u}_t^*$ como

$$\hat{u}_t^* = Y_t - \hat{\beta}_1^* - \hat{\beta}_2^* X_t \tag{4}$$

que se calculan con facilidad, pues se conocen $Y_t$, $X_t$, $\hat{\beta}_1^*$ y $\hat{\beta}_2^*$.

5. Ahora calcule la siguiente regresión:

$$\hat{u}_t^* = \hat{\rho}^* \hat{u}_{t-1}^* + w_t \tag{5}$$

que es similar a (3), y por tanto proporciona el estimado de $\rho$ de la segunda ronda.

Como desconocemos si dicha estimación de $\rho$ es la mejor estimación de la verdadera $\rho$, calculamos la estimación de la tercera ronda, y así sucesivamente. Por esta razón el procedimiento C-O se llama método iterativo. Pero, ¿hasta dónde continuamos iterando? La recomendación general es detener las iteraciones cuando las estimaciones sucesivas de $\rho$ difieran por una pequeña cantidad, por ejemplo, menores que 0.01 o 0.005. En el ejemplo de la regresión de los salarios sobre la productividad se requirieron alrededor de tres iteraciones antes de detenerse.

*a*) Con el procedimiento iterativo de Cochrane-Orcutt estime $\rho$ en la regresión de los salarios sobre la productividad, ecuación (12.5.2). ¿Cuántas iteraciones se requirieron para obtener la estimación "final" de $\rho$?

*b*) Con la estimación final de $\rho$ obtenida en *a*), estime la regresión de los salarios sobre la productividad, tanto eliminando la primera observación como conservándola. ¿Qué diferencia observa en los resultados?

*c*) ¿Cree que es importante conservar la primera observación para transformar los datos con el fin de resolver el problema de la autocorrelación?

12.9. *Estimación de ρ: procedimiento de dos pasos de Cochrane-Orcutt.* Es una versión abreviada del procedimiento iterativo C-O. En el paso 1 se estima $\rho$ a partir de la primera iteración, es decir, de la ecuación (3) del ejercicio anterior, y en el paso 2 se utiliza la estimación de $\rho$ para efectuar la ecuación en diferencias generalizada, como en la ecuación (4) del ejercicio anterior. A veces en la práctica este método de dos pasos proporciona resultados muy similares a los del procedimiento iterativo C-O, más elaborado.

Aplique el método de dos pasos C-O para ilustrar la regresión de los salarios sobre la productividad (12.5.1) de este capítulo y compare los resultados con los obtenidos mediante el método iterativo. Ponga especial atención a la primera observación en la transformación.

12.10. *Estimación de ρ: método de dos pasos de Durbin.** Para explicar este método, expresamos de forma equivalente la ecuación en diferencia generalizada (12.9.5) como:

$$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + \varepsilon_t \tag{1}$$

Durbin propone el siguiente procedimiento de dos pasos para calcular $\rho$. *Primero*, considere (1) como un modelo de regresión múltiple, haga la regresión $Y_t$ sobre $X_t$, $X_{t-1}$ y $Y_{t-1}$, y considere el valor estimado del coeficiente de la regresión de $Y_{t-1}$ $(= \hat{\rho})$ como una estimación de $\rho$. *Segundo*, tras obtener $\hat{\rho}$, utilícelo para estimar los parámetros de la ecuación en diferencias generalizada (12.9.5) o su equivalente (12.9.6).

* J. Durbin, "Estimation of Parameters in Time Series Regression Models", *Journal of the Royal Statistical Society*, serie B, vol. 22, 1960, pp. 139-153.

*a*) Aplique el método de dos pasos de Durbin al ejemplo de la regresión de los salarios sobre la productividad, analizado antes en el libro, y compare los resultados con los del procedimiento iterativo Cochrane-Orcutt y los del método de dos pasos C-O. Asimismo, comente respecto de la "calidad" de sus resultados.

*b*) Si examina la ecuación (1) inmediata anterior, observará que el coeficiente de $X_{t-1}$ $(= -\rho\beta_2)$ es igual a menos 1 por el producto del coeficiente de $X_t (= \beta_2)$ y el coeficiente de $Y_{t-1}$ $(= \rho)$. ¿Cómo probaría que los coeficientes obedecen la restricción anterior?

12.11. Al medir los rendimientos a escala en la oferta de electricidad, Nerlove utilizó información de corte transversal de 145 empresas de servicios de propiedad privada en Estados Unidos durante 1955 y efectuó la regresión del logaritmo del costo total sobre los logaritmos de la producción, de la tasa de salarios, del precio del capital y del precio del combustible. Encontró que los residuos estimados a partir de esta regresión presentaban correlación "serial" a juicio del *d* de Durbin-Watson. Para remediarlo, graficó los residuos estimados respecto del logaritmo de la producción y obtuvo la figura 12.11.

*a*) ¿Qué indica la figura 12.11?

*b*) ¿Cómo puede eliminar la correlación "serial" en la situación anterior?

12.12. Al graficar los residuos de una regresión respecto del tiempo, se obtuvo el diagrama de dispersión de la figura 12.12. El residuo "extremo" encerrado en un círculo se denomina *valor atípico*. Un valor atípico es una observación cuyo valor excede los valores de las

**FIGURA 12.11**
Residuos de regresión del estudio de Nerlove. (Adaptado de Marc Nerlove, "Return to Scale in Electric Supply", en Carl F. Christ *et al.*, *Measurement in Economics*, Stanford University Press, Stanford, California, 1963.)

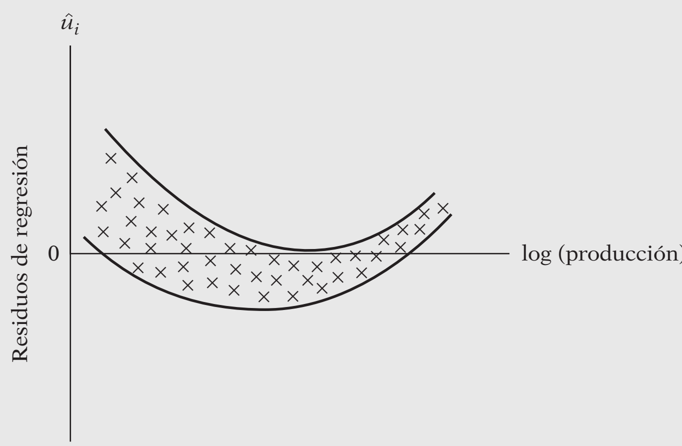
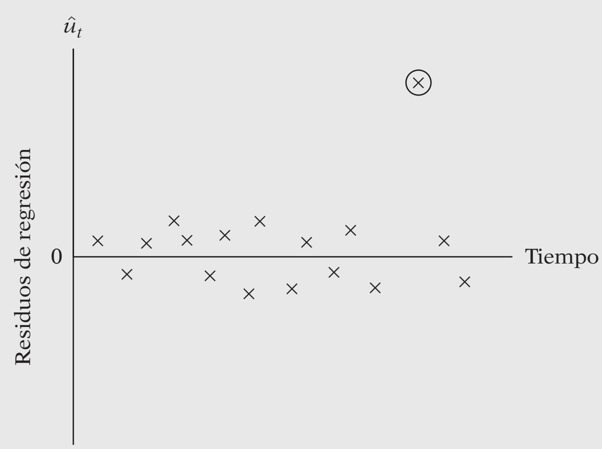



**FIGURA 12.12**
Residuos de regresión hipotéticos graficados respecto del tiempo.



demás observaciones en la muestra por una gran cantidad, tal vez tres o cuatro desviaciones estándar alejada del valor medio de todas las observaciones.

*a*) ¿Cuáles son las razones de la existencia de uno o varios valores atípicos?

*b*) Si hay uno o varios valores atípicos, ¿deben descartarse esas observaciones y efectuar la regresión sobre las observaciones restantes?

*c*) ¿Es aplicable el $d$ de Durbin-Watson en presencia de valores atípicos?

12.13. Con base en el estadístico $d$ de Durbin-Watson, ¿cómo distinguiría la autocorrelación "pura" del sesgo de especificación?

12.14. Suponga que en el modelo

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

las $u$ son en realidad serialmente independientes. ¿Qué sucedería en esta situación si, suponiendo que $u_t = \rho u_{t-1} + \varepsilon_t$, utilizáramos la siguiente regresión en diferencia generalizada?

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \varepsilon_t$$

Analice en particular las propiedades del término de perturbación $\varepsilon_t$.

12.15. En un estudio de determinación de precios de la producción final a costo de factor en el Reino Unido se obtuvieron los siguientes resultados con base en los datos anuales de 1951 a 1969:

$$\begin{aligned}\widehat{\text{PF}}_t = {} & 2.033 + 0.273W_t - 0.521X_t + 0.256M_t + 0.028M_{t-1} + 0.121PF_{t-1}\\ \text{ee} = {} & (0.992) \quad (0.127) \quad (0.099) \quad (0.024) \quad (0.039) \quad (0.119)\\ & R^2 = 0.984 \qquad d = 2.54\end{aligned}$$

donde PF = precios de la producción final a costo de factor, $W$ = salarios por empleado, $X$ = producto interno bruto por persona empleada, $M$ = precios de importación, $M_{t-1}$ = precios de importación rezagados 1 año y $\text{PF}_{t-1}$ = precios de la producción final a costo de factor en el año anterior.*

"Como para 18 observaciones y 5 variables explicativas a 5% los valores $d$ inferior y superior son 0.71 y 2.06, el valor $d$ estimado de 2.54 indica que no hay autocorrelación positiva". Comente.

12.16. Establezca las circunstancias en que sería adecuado cada uno de los siguientes métodos de estimación del coeficiente de autocorrelación de primer orden $\rho$:

*a*) Regresión de primeras diferencias

*b*) Regresión de promedios móviles

*c*) Transformación Theil-Nagar

*d*) Procedimiento iterativo Cochrane y Orcutt

*e*) Procedimiento de exploración Hildreth-Lu

*f*) Procedimiento en dos etapas de Durbin

12.17. Considere el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

donde

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

* Fuente: *Prices and Earnings in 1951-1969: An Econometric Assessment*, Department of Employment, Her Majesty's Stationery Office, 1971, tabla C, p. 37, ecuación 63.

es decir, el término de error sigue un esquema AR(2), y $\varepsilon_t$ es un término de error de ruido blanco. Describa los pasos que seguiría para estimar el modelo si tiene en cuenta la autorregresión de segundo orden.

12.18. Con el factor de corrección C, la fórmula para $\hat{\beta}_2^{\text{MCG}}$ dada en (12.3.1) es

$$\hat{\beta}_2^{\text{MCG}} = \frac{(1-\rho^2)x_1y_1 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{(1-\rho^2)x_1^2 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})^2}$$

Con esta fórmula y (12.3.1), encuentre la expresión para el factor de corrección $C$.

12.19. Muestre que la estimación de (12.9.5) equivale a estimar los MCG analizados en la sección 12.3, excluyendo la primera observación en $Y$ y en $X$.

12.20. Para la regresión (12.9.9), los residuos estimados tuvieron los siguientes signos:

$$(++++)(-)(+++++++)(-)(++++)(--)(+)(--)(+)(--)(++)(-)$$
$$(+)(---------)(+)$$

Con base en la prueba de rachas, ¿rechaza la hipótesis nula de que no hay autocorrelación en estos residuos?

*12.21. *Prueba para correlación serial de orden superior*. Suponga que tenemos información de series de tiempo sobre una base trimestral. En los modelos de regresión que consideran información trimestral, en lugar de utilizar el esquema AR(1) dado en (12.2.1), puede ser más apropiado suponer un esquema AR(4) como el siguiente:

$$u_t = \rho_4 u_{t-4} + \varepsilon_t$$

es decir, suponer que el término de perturbación actual está correlacionado con el término para el mismo trimestre del año anterior, en lugar de estarlo con el del trimestre anterior.

Para probar la hipótesis de que $\rho_4 = 0$, Wallis[†] sugiere la siguiente prueba $d$ modificada de Durbin-Watson:

$$d_4 = \frac{\sum_{t=5}^{n}(\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

El procedimiento de prueba sigue la rutina de la prueba $d$ usual analizada en el texto. Wallis preparó las tablas $d_4$, las cuales se encuentran en su artículo original.

Suponga ahora que hay información mensual. ¿Puede generalizarse la prueba Durbin-Watson para considerar tal información? De ser así, escriba la fórmula $d_{12}$ adecuada.

12.22. Suponga que estima la siguiente regresión:

$$\Delta \ln \text{producción}_t = \beta_1 + \beta_2 \Delta \ln L_t + \beta_3 \Delta \ln K_t + u_t$$

donde $Y$ es la producción, $L$ es el insumo trabajo, $K$ es el insumo capital y $\Delta$ es el operador de primeras diferencias. ¿Cómo interpretaría $\beta_1$ en este modelo? ¿Puede verse como una estimación del cambio tecnológico? Justifique su respuesta.

---

* Opcional.

[†] Kenneth Wallis, "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations", *Econometrica*, vol. 40, 1972, pp. 617-636. Las tablas de $d_4$ se encuentran también en J. Johnston, *op.cit.,* 3a. ed., p. 558.

12.23. Maddala sugiere que si el $d$ de Durbin-Watson es menor que $R^2$, se debe efectuar la regresión en forma de primeras diferencias. ¿Cuál es el razonamiento de esta sugerencia?

12.24. Consulte la ecuación (12.4.1). Suponga $r = 0$, pero $\rho \neq 0$. ¿Cuál es el efecto sobre $E(\hat{\sigma}^2)$ si *a*) $0 < \rho < 1$ y *b*) $-1 < \rho < 0$? ¿Cuándo será razonablemente pequeño el sesgo en $\hat{\sigma}^2$?

12.25. Se hizo la regresión de los residuos de la regresión de los salarios sobre la productividad dados en (12.5.2), sobre los residuos rezagados de seis periodos anteriores [es decir, AR(6)], con los siguientes resultados:

```
Variable dependiente: S1
Método: Mínimos cuadrados
Muestra (ajustada): 1966-2005
Observaciones incluidas: 40 después de los ajustes
```

| | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| SI(-1) | 1.019716 | 0.170999 | -5.963275 | 0.0000 |
| SI(-2) | -0.029679 | 0.244152 | -0.121560 | 0.9040 |
| SI(-3) | -0.286782 | 0.241975 | -1.185171 | 0.2442 |
| SI(-4) | 0.149212 | 0.242076 | 0.616386 | 0.5417 |
| SI(-5) | -0.071371 | 0.243386 | -0.293240 | 0.7711 |
| SI(-6) | 0.034362 | 0.167077 | 0.205663 | 0.8383 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.749857 | Media de la variable dependiente | 7.004433 |
| R cuadrada ajustada | 0.713071 | Desviación estándar de la variable dependiente | 0.019843 |
| Error estándar de la regresión | 0.010629 | | |
| Suma de cuadrados residual | 0.003741 | Estadístico de Durbin-Watson | 1.956818 |

*a*) De estos resultados, ¿qué puede decir sobre la naturaleza de la autocorrelación en los datos sobre salarios y productividad?

*b*) Si piensa que un mecanismo AR(1) caracteriza la autocorrelación en los datos, ¿utilizaría la transformación de primeras diferencias para eliminar la autocorrelación? Justifique su respuesta.

## Ejercicios empíricos

12.26. Consulte los datos sobre la industria del cobre de la tabla 12.7.

*a*) Con base en esta información, estime el siguiente modelo de regresión:

$$\ln C_t = \beta_1 + \beta_2 \ln I_t + \beta_3 \ln L_t + \beta_4 \ln H_t + \beta_5 \ln A_t + u_t$$

Interprete los resultados.

*b*) Obtenga los residuos y los residuos estandarizados de la regresión anterior y grafíquelos. ¿Qué opina sobre la presencia de autocorrelación en estos residuos?

*c*) Estime el estadístico $d$ de Durbin-Watson y comente sobre la naturaleza de la autocorrelación presente en los datos.

*d*) Efectúe la prueba de rachas y vea si su respuesta difiere de la respuesta dada en *c*).

*e*) ¿Cómo investigaría si un proceso AR(p) describe mejor la autocorrelación que un proceso AR(1)?

*Nota:* guarde estos datos para un análisis posterior (ejercicio 12.28).

**TABLA 12.7**
**Determinantes del precio interno del cobre en Estados Unidos, 1951-1980**

| Año | *C* | *G* | *I* | *L* | *H* | *A* |
|---|---|---|---|---|---|---|
| 1951 | 21.89 | 330.2 | 45.1 | 220.4 | 1 491.0 | 19.00 |
| 52 | 22.29 | 347.2 | 50.9 | 259.5 | 1 504.0 | 19.41 |
| 53 | 19.63 | 366.1 | 53.3 | 256.3 | 1 438.0 | 20.93 |
| 54 | 22.85 | 366.3 | 53.6 | 249.3 | 1 551.0 | 21.78 |
| 55 | 33.77 | 399.3 | 54.6 | 352.3 | 1 646.0 | 23.68 |
| 56 | 39.18 | 420.7 | 61.1 | 329.1 | 1 349.0 | 26.01 |
| 57 | 30.58 | 442.0 | 61.9 | 219.6 | 1 224.0 | 27.52 |
| 58 | 26.30 | 447.0 | 57.9 | 234.8 | 1 382.0 | 26.89 |
| 59 | 30.70 | 483.0 | 64.8 | 237.4 | 1 553.7 | 26.85 |
| 60 | 32.10 | 506.0 | 66.2 | 245.8 | 1 296.1 | 27.23 |
| 61 | 30.00 | 523.3 | 66.7 | 229.2 | 1 365.0 | 25.46 |
| 62 | 30.80 | 563.8 | 72.2 | 233.9 | 1 492.5 | 23.88 |
| 63 | 30.80 | 594.7 | 76.5 | 234.2 | 1 634.9 | 22.62 |
| 64 | 32.60 | 635.7 | 81.7 | 347.0 | 1 561.0 | 23.72 |
| 65 | 35.40 | 688.1 | 89.8 | 468.1 | 1 509.7 | 24.50 |
| 66 | 36.60 | 753.0 | 97.8 | 555.0 | 1 195.8 | 24.50 |
| 67 | 38.60 | 796.3 | 100.0 | 418.0 | 1 321.9 | 24.98 |
| 68 | 42.20 | 868.5 | 106.3 | 525.2 | 1 545.4 | 25.58 |
| 69 | 47.90 | 935.5 | 111.1 | 620.7 | 1 499.5 | 27.18 |
| 70 | 58.20 | 982.4 | 107.8 | 588.6 | 1 469.0 | 28.72 |
| 71 | 52.00 | 1 063.4 | 109.6 | 444.4 | 2 084.5 | 29.00 |
| 72 | 51.20 | 1 171.1 | 119.7 | 427.8 | 2 378.5 | 26.67 |
| 73 | 59.50 | 1 306.6 | 129.8 | 727.1 | 2 057.5 | 25.33 |
| 74 | 77.30 | 1 412.9 | 129.3 | 877.6 | 1 352.5 | 34.06 |
| 75 | 64.20 | 1 528.8 | 117.8 | 556.6 | 1 171.4 | 39.79 |
| 76 | 69.60 | 1 700.1 | 129.8 | 780.6 | 1 547.6 | 44.49 |
| 77 | 66.80 | 1 887.2 | 137.1 | 750.7 | 1 989.8 | 51.23 |
| 78 | 66.50 | 2 127.6 | 145.2 | 709.8 | 2 023.3 | 54.42 |
| 79 | 98.30 | 2 628.8 | 152.5 | 935.7 | 1 749.2 | 61.01 |
| 80 | 101.40 | 2 633.1 | 147.1 | 940.9 | 1 298.5 | 70.87 |

*Nota:* Gary R. Smith recopiló los datos de fuentes como *American Metal Market, Metals Week* y publicaciones del Departamento de Comercio de Estados Unidos.
$C$ = promedio de doce meses del precio interno del cobre en Estados Unidos (centavos por libra).
$G$ = Producto Nacional Bruto anual (miles de millones).
$I$ = Índice promedio de doce meses de la producción industrial.
$L$ = Precio promedio de doce meses del cobre en la bolsa de metales de Londres (libras esterlinas).
$H$ = Número de casas construidas por año (miles de unidades).
$A$ = Precio promedio de doce meses del aluminio (centavos de dólar por libra).

12.27. Se proporcionan los datos de la tabla 12.8.

*a*) Verifique que el $d$ de Durbin-Watson es igual a 0.4148.

*b*) ¿Hay correlación serial positiva en las perturbaciones?

*c*) De ser así, estime $\rho$ mediante el

*i*) Método de Theil-Nagar.

*ii*) Procedimiento de dos pasos de Durbin.

*iii*) Método de Cochrane-Orcutt.

*d*) Con el método de Theil-Nagar transforme los datos y efectúe la regresión con los datos transformados.

*e*) ¿La regresión estimada en *d*) presenta autocorrelación? De ser así, ¿cómo se desharía de ella?

**TABLA 12.8**

| $Y$, gasto de consumo personal, miles de millones de dólares de 1958 | $X$, tiempo | $\hat{Y}$, $Y$ estimado | $\hat{u}$, residuos |
|---|---|---|---|
| 281.4 | 1 (= 1956) | 261.4208 | 19.9791 |
| 288.1 | 2 | 276.6026 | 11.4973 |
| 290.0 | 3 | 291.7844 | −1.7844 |
| 307.3 | 4 | 306.9661 | 0.3338 |
| 316.1 | 5 | 322.1479 | −6.0479 |
| 322.5 | 6 | 337.3297 | −14.8297 |
| 338.4 | 7 | 352.5115 | −14.1115 |
| 353.3 | 8 | 367.6933 | −14.3933 |
| 373.7 | 9 | 382.8751 | −9.1751 |
| 397.7 | 10 | 398.0569 | −0.3569 |
| 418.1 | 11 | 413.2386 | 4.8613 |
| 430.1 | 12 | 428.4206 | 1.6795 |
| 452.7 | 13 | 443.6022 | 9.0977 |
| 469.1 | 14 | 458.7840 | 10.3159 |
| 476.9 | 15 (= 1970) | 473.9658 | 2.9341 |

*Nota:* Datos de $\hat{Y}$ obtenidos de la regresión $Y_t = \beta_0 + \beta_1 X_t + u_t$.

12.28. Consulte el ejercicio 12.26 y los datos de la tabla 12.7. Si los resultados de este ejercicio muestran correlación serial,

*a*) Utilice el procedimiento de Cochrane-Orcutt de dos etapas y obtenga las estimaciones de MCG factibles o la regresión en diferencias generalizada, y compare sus resultados.

*b*) Si el $\rho$ estimado del método de Cochrane-Orcutt en *a*) difiere sustancialmente del estimado a partir del estadístico $d$, ¿qué método de estimación de $\rho$ seleccionaría y por qué?

12.29. Consulte el ejemplo 7.4. Efectúe la regresión omitiendo las variables $X^2$ y $X^3$, y examine los residuos en busca de correlación "serial". Si encuentra correlación serial, ¿cómo se explicaría su presencia? ¿Qué medidas correctivas recomendaría?

12.30. Consulte el ejercicio 7.21. Con tal información, *a priori* se espera autocorrelación. Por consiguiente, se sugiere que efectúe la regresión del logaritmo de la oferta monetaria real sobre los logaritmos del ingreso nacional real y de la tasa de interés de largo plazo en la forma de primeras diferencias. Efectúe esta regresión y luego hágala de nuevo en la forma original. ¿Se satisface el supuesto en que se basa la transformación de primeras diferencias? De no ser así, ¿qué clases de sesgos es probable que resulten de tal transformación? Explique su respuesta con los datos disponibles.

12.31. *Uso del* d *de Durbin-Watson para probar la no linealidad.* Continúe con el ejercicio 12.29. Ordene los residuos obtenidos en esa regresión de acuerdo con valores crecientes de $X$. Con la fórmula dada en (12.6.5), estime $d$ a partir de los residuos reordenados. Si el valor $d$ calculado indica autocorrelación, esto implicaría que el modelo lineal era incorrecto y que el modelo completo debe incluir los términos $X_i^2$ y $X_i^3$. ¿Puede dar una justificación intuitiva para tal procedimiento? Vea si la respuesta coincide con la de Henri Theil.*

12.32. Consulte el ejercicio 11.22. Obtenga los residuos y vea si hay autocorrelación en ellos. ¿Cómo transformaría los datos en caso de detectar correlación serial? ¿Cuál es el significado de la correlación serial en el presente ejemplo?

* Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, pp. 307-308.

12.33. *Experimento Monte Carlo.* Consulte las tablas 12.1 y 12.2. Con los datos sobre $e_t$ y $X_i$ que allí se presentan, genere una muestra de 10 valores $Y$ del modelo

$$Y_t = 3.0 + 0.5X_t + u_t$$

donde $u_t = 0.9u_{t-1} + \varepsilon_t$. Suponga que $u_0 = 10$.

*a*) Calcule la ecuación. Comente sus resultados.

*b*) Ahora suponga su $u_0 = 17$. Repita este ejercicio 10 veces y comente sus resultados.

*c*) Mantenga intactos los datos anteriores, excepto que ahora permita que $\rho = 0.3$ en lugar de $\rho = 0.9$ y compare sus resultados con los obtenidos en *b*).

12.34. Con los datos proporcionados en la tabla 12.9, estime el modelo

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

donde $Y$ = inventarios y $X$ = ventas, ambas medidas en miles de millones de dólares.

*a*) Estime la regresión anterior.

*b*) Con los residuos estimados, investigue si hay autocorrelación positiva mediante *i*) la prueba de Durbin-Watson y *ii*) la prueba de normalidad para grandes muestras dada en (12.6.13).

*c*) Si $\rho$ es positivo, aplique la prueba de Berenblutt-Webb para evaluar la hipótesis de que $\rho = 1$.

*d*) Si sospecha que la estructura autorregresiva del error es de orden $p$, verifíquelo con la prueba de Breusch-Godfrey. ¿Cómo seleccionaría el orden de $p$?

*e*) Con base en los resultados de esta prueba, ¿cómo transformaría los datos para eliminar la autocorrelación? Muestre todos sus cálculos.

**TABLA 12.9** **Inventarios y ventas en la industria manufacturera de Estados Unidos, 1950-1991 (millones de dólares)**

| Año | Ventas* | Inventarios† | Razón | Año | Ventas* | Inventarios† | Razón |
|---|---|---|---|---|---|---|---|
| 1950 | 46 486 | 84 646 | 1.82 | 1971 | 224 619 | 369 374 | 1.57 |
| 1951 | 50 229 | 90 560 | 1.80 | 1972 | 236 698 | 391 212 | 1.63 |
| 1952 | 53 501 | 98 145 | 1.83 | 1973 | 242 686 | 405 073 | 1.65 |
| 1953 | 52 805 | 101 599 | 1.92 | 1974 | 239 847 | 390 950 | 1.65 |
| 1954 | 55 906 | 102 567 | 1.83 | 1975 | 250 394 | 382 510 | 1.54 |
| 1955 | 63 027 | 108 121 | 1.72 | 1976 | 242 002 | 378 762 | 1.57 |
| 1956 | 72 931 | 124 499 | 1.71 | 1977 | 251 708 | 379 706 | 1.50 |
| 1957 | 84 790 | 157 625 | 1.86 | 1978 | 269 843 | 399 970 | 1.44 |
| 1958 | 86 589 | 159 708 | 1.84 | 1979 | 289 973 | 424 843 | 1.44 |
| 1959 | 98 797 | 174 636 | 1.77 | 1980 | 299 766 | 430 518 | 1.43 |
| 1960 | 113 201 | 188 378 | 1.66 | 1981 | 319 558 | 443 622 | 1.37 |
| 1961 | 126 905 | 211 691 | 1.67 | 1982 | 324 984 | 449 083 | 1.38 |
| 1962 | 143 936 | 242 157 | 1.68 | 1983 | 335 991 | 463 563 | 1.35 |
| 1963 | 154 391 | 265 215 | 1.72 | 1984 | 350 715 | 481 633 | 1.35 |
| 1964 | 168 129 | 283 413 | 1.69 | 1985 | 330 875 | 428 108 | 1.38 |
| 1965 | 163 351 | 311 852 | 1.95 | 1986 | 326 227 | 423 082 | 1.29 |
| 1966 | 172 547 | 312 379 | 1.78 | 1987 | 334 616 | 408 226 | 1.24 |
| 1967 | 190 682 | 339 516 | 1.73 | 1988 | 359 081 | 439 821 | 1.18 |
| 1968 | 194 538 | 334 749 | 1.73 | 1989 | 394 615 | 479 106 | 1.17 |
| 1969 | 194 657 | 322 654 | 1.68 | 1990 | 411 663 | 509 902 | 1.21 |
| 1970 | 206 326 | 338 109 | 1.59 | | | | |

* Los datos anuales son promedios de cifras mensuales sin ajustar por estacionalidad.

† Las cifras de fin de periodo, ajustadas por estacionalidad, comienzan en 1982 y no son comparables con periodos anteriores.

Fuente: *Economic Report of the President*, 1993, tabla B-53, p. 408.

*f*) Repita los pasos anteriores con el siguiente modelo:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_t + u_t$$

*g*) ¿Cómo decidiría entre la especificación lineal y la log-lineal? Muestre explícitamente la(s) prueba(s) que utiliza.

12.35. La tabla 12.10 proporciona datos sobre la tasa real de rendimiento para acciones comunes en el tiempo *t* ($RR_t$), el crecimiento de la producción en el periodo $(t + 1)$ ($CP_{t+1}$) y la inflación en el periodo *t* ($Inf_t$), todos dados en porcentaje, en la economía estadounidense de 1954 a 1981.

*a*) Haga la regresión de $RR_t$ sobre la inflación.

*b*) Haga la regresión de $RR_t$ sobre el $CP_{t+1}$ e $Inf_t$.

*c*) Comente los resultados de estas dos regresiones a la luz del comentario de Eugene Fama, respecto de que "la correlación simple negativa entre los rendimientos reales de las acciones y la inflación es espuria porque es el resultado de dos relaciones estructurales: una relación positiva entre los rendimientos reales actuales de las acciones y el crecimiento de la producción esperado [medido por $CP_{t+1}$], y una relación negativa entre el crecimiento de la producción esperado y la inflación actual".

*d*) ¿Esperaría autocorrelación en cualquiera de las dos regresiones de *a*) y *b*)? ¿Por qué? Si existiera, tome las medidas correctivas apropiadas y presente los resultados revisados.

**TABLA 12.10**
**Tasa de rendimiento, crecimiento de la producción e inflación, Estados Unidos, 1954-1981**

| Observación | RR | Crecimiento | Inflación |
|---|---|---|---|
| 1954 | 53.0 | 6.7 | –0.4 |
| 1955 | 31.2 | 2.1 | 0.4 |
| 1956 | 3.7 | 1.8 | 2.9 |
| 1957 | –13.8 | –0.4 | 3.0 |
| 1958 | 41.7 | 6.0 | 1.7 |
| 1959 | 10.5 | 2.1 | 1.5 |
| 1960 | –1.3 | 2.6 | 1.8 |
| 1961 | 26.1 | 5.8 | 0.8 |
| 1962 | –10.5 | 4.0 | 1.8 |
| 1963 | 21.2 | 5.3 | 1.6 |
| 1964 | 15.5 | 6.0 | 1.0 |
| 1965 | 10.2 | 6.0 | 2.3 |
| 1966 | –13.3 | 2.7 | 3.2 |
| 1967 | 21.3 | 4.6 | 2.7 |
| 1968 | 6.8 | 2.8 | 4.3 |
| 1969 | –13.5 | –0.2 | 5.0 |
| 1970 | –0.4 | 3.4 | 4.4 |
| 1971 | 10.5 | 5.7 | 3.8 |
| 1972 | 15.4 | 5.8 | 3.6 |
| 1973 | –22.6 | –0.6 | 7.9 |
| 1974 | –37.3 | –1.2 | 10.8 |
| 1975 | 31.2 | 5.4 | 6.0 |
| 1976 | 19.1 | 5.5 | 4.7 |
| 1977 | –13.1 | 5.0 | 5.9 |
| 1978 | –1.3 | 2.8 | 7.9 |
| 1979 | 8.6 | –0.3 | 9.8 |
| 1980 | –22.2 | 2.6 | 10.2 |
| 1981 | –12.2 | –1.9 | 7.3 |

12.36. *Estadístico h de Durbin.* Considere el siguiente modelo de la determinación de salarios:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

donde $Y$ = salarios = índice de remuneración real por hora
$X$ = productividad = índice de producción por hora

*a*) Con los datos de la tabla 12.4, calcule el modelo anterior e interprete sus resultados.

*b*) Como el modelo contiene a la regresada rezagada como variable regresora, el $d$ de Durbin-Watson no resulta apropiado para averiguar si existe correlación serial en los datos. Para tales modelos, llamados autorregresivos, Durbin desarrolló el **estadístico *h*** para probar la autocorrelación de primer orden, el cual se define como:*

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\text{var}(\hat{\beta}_3)]}}$$

donde $n$ = tamaño de la muestra, var($\hat{\beta}_3$) = varianza del coeficiente de la $Y_{t-1}$ rezagada y $\hat{\rho}$ = la estimación de la correlación serial de primer orden.

Para un tamaño de muestra grande (técnicamente asintótica), Durbin mostró que, según la hipótesis nula de que $\rho = 0$,

$$h \sim N(0, 1)$$

es decir, el estadístico $h$ sigue la distribución normal estandarizada. A partir de las propiedades de la distribución normal, sabemos que la probabilidad de que $|h| > 1.96$ es de casi 5%. Por consiguiente, si en una aplicación $|h| > 1.96$, podemos rechazar la hipótesis nula de que $\rho = 0$; es decir, existe evidencia de autocorrelación de primer orden en el modelo autorregresivo dado antes.

Para aplicar la prueba procedemos así: *primero* estime el modelo anterior mediante MCO (en este momento no hay que preocuparse por problemas de estimación). *Segundo*, observe var($\hat{\beta}_3$) en este modelo, así como el estadístico $d$ que se calcula de manera rutinaria. *Tercero*, con el valor $d$, obtenga $\hat{\rho} \approx (1 - d/2)$. Resulta interesante notar que, a pesar de que no se puede emplear el valor $d$ para probar la correlación serial en este modelo, sí se puede usar para obtener una estimación de $\rho$. *Cuarto*, ahora calcule el estadístico $h$. *Quinto*, si el tamaño de la muestra es razonablemente grande y la $|h|$ calculada excede 1.96, podemos concluir que hay evidencia de autocorrelación de primer orden. Por supuesto, puede usar cualquier nivel de significancia que desee.

Aplique la prueba $h$ al modelo autorregresivo de determinación del salario dado antes y deduzca las conclusiones apropiadas. También compare los resultados con los obtenidos mediante la regresión (12.5.1).

12.37. *Variables dicótomas y autocorrelación.* Consulte la regresión de los ahorros sobre el ingreso analizada en el capítulo 9. Con los datos de la tabla 9.2 y suponiendo un esquema AR(1), vuelva a realizar la regresión de los ahorros sobre el ingreso, tomando en cuenta la autocorrelación. Ponga especial atención a la transformación de la variable dicótoma. Compare sus resultados con los del capítulo 9.

12.38. Con los datos para la regresión de los salarios sobre la productividad de la tabla 12.4, estime el modelo (12.9.8) y compare los resultados con los obtenidos mediante la regresión (12.9.9). ¿Qué conclusión(es) deduce?

---

* J. Durbin, "Testing for Serial Correlation in Least-squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, vol. 38, pp. 410-421

# Apéndice 12A

## 12A.1 Prueba de que el término de error $v_t$ en la ecuación (12.1.11) está autocorrelacionado

Como $v_t = u_t - u_{t-1}$, resulta fácil demostrar que $E(v_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$, pues $E(u) = 0$, para cada $t$. Ahora bien, $\text{var}(v_t) = \text{var}(u_t - u_{t-1}) = \text{var}(u_t) + \text{var}(u_{t-1}) = 2\sigma^2$, pues la varianza de cada $u_t$ es $\sigma^2$ y las $u$ están independientemente distribuidas. Por tanto, $v_t$ es homoscedástica. Pero

$$\begin{aligned}\text{cov}(v_t, v_{t-1}) &= E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] \\ &= -\sigma^2\end{aligned}$$

lo cual es obviamente diferente de cero. En consecuencia, aunque las $u$ no están autocorrelacionadas, las $v$ sí lo están.

## 12A.2 Pruebas de las ecuaciones (12.2.3), (12.2.4) y (12.2.5)

Con AR(1),

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{1}$$

Por tanto,

$$E(u_t) = \rho E(u_{t-1}) + E(\varepsilon_t) = 0 \tag{2}$$

Así que,

$$\text{var}(u_t) = \rho^2 \text{ var}(u_{t-1}) + \text{var}(\varepsilon_t) \tag{3}$$

porque las $u$ y las $\varepsilon$ no están correlacionados.

Como $\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma^2$ y $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$, tenemos

$$\text{var}(u_t) = \frac{\sigma_\varepsilon^2}{1 - \rho^2} \tag{4}$$

Ahora multiplicamos (1) por $u_{t-1}$ y tomamos las esperanzas en ambos miembros para obtener:

$$\text{cov}(u_t, u_{t-1}) = E(u_t u_{t-1}) = E\left[\rho u_{t-1}^2 + u_{t-1}\varepsilon_t\right] = \rho E\left(u_{t-1}^2\right)$$

Se observa que la covarianza entre $u_{t-1}$ y $\varepsilon_t$ es cero (¿por qué?), y además que $\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma_\varepsilon^2/(1 - \rho^2)$, y obtenemos

$$\text{cov}(u_t, u_{t-1}) = \rho \frac{\sigma_\varepsilon^2}{(1 - \rho^2)} \tag{5}$$

Para continuar de la misma forma,

$$\text{cov}(u_t, u_{t-2}) = \rho^2 \frac{\sigma_\varepsilon^2}{(1 - \rho^2)}$$

$$\text{cov}(u_t, u_{t-3}) = \rho^3 \frac{\sigma_\varepsilon^2}{(1 - \rho^2)}$$

y así sucesivamente. Ahora el coeficiente de correlación es la razón de la covarianza respecto de la varianza. Por tanto,

$$\text{cor}(u_t, u_{t-1}) = \rho \qquad \text{cov}(u_t, u_{t-2}) = \rho^2$$

y así sucesivamente.

Capítulo 13

# Creación de modelos econométricos: especificación del modelo y pruebas de diagnóstico

> La econometría aplicada no puede concebirse de manera mecánica: necesita comprensión, intuición y habilidades.[1]

> . . . por lo general atravesamos puentes sin preocuparnos por la solidez de su construcción, pues tenemos la certeza razonable de que alguien verificó con rigor los principios y prácticas de la ingeniería. Los economistas deben hacer esta verificación con los modelos, o al menos anexar la siguiente advertencia a su modelo: "no nos hacemos responsables si al emplearse se provoca un colapso".[2]

> La búsqueda de la "verdad" por parte de los economistas a lo largo de los años ha dado origen al punto de vista según el cual los economistas son personas que buscan en un cuarto oscuro un gato negro que no existe; a los econometristas por lo general se les acusa de haberlo encontrado.[3]

Un supuesto del modelo clásico de regresión lineal (MCRL), el 9, es que el modelo de regresión del análisis está especificado "correctamente"; si no es así, nos enfrentamos al problema de **error de especificación del modelo** o **sesgo en la especificación del modelo**. En este capítulo revisaremos con una mirada cuidadosa y crítica este supuesto, pues la búsqueda del modelo correcto se asemeja a la del Santo Grial. En particular, examinaremos las siguientes preguntas:

1. ¿Qué hacer para encontrar el modelo "correcto"? En otras palabras, ¿cuáles son los criterios para elegir un modelo a partir del análisis empírico?
2. ¿Qué tipos de errores de especificación de modelos son más comunes en la práctica?
3. ¿Cuáles son las consecuencias de los errores de especificación?
4. ¿Cómo se detectan los errores de especificación? En otras palabras, ¿cuáles son algunas herramientas de diagnóstico disponibles?
5. Una vez detectados los errores de especificación, ¿cuáles son los remedios que se pueden adoptar y con qué beneficios?
6. ¿Cómo se evalúa el desempeño de modelos contendientes?

[1] Keith Cuthberston, Stephen G. Hall y Mark P. Taylor, *Applied Econometrics Techniques,* Michigan University Press, 1992, p. X.

[2] David F. Hendry, *Dynamic Econometrics,* Oxford University Press, Inglaterra, 1995, p. 68.

[3] Peter Kennedy, *A Guide to Econometrics,* 3a. ed., The MIT Press, Cambridge Massachusetts, 1992, p. 82.

El tema de la especificación y evaluación de modelos es amplio, así que se requiere una tarea empírica muy extensa en esta área. Y no sólo eso, sino que también existen diferencias filosóficas sobre este asunto. Aunque no podemos abarcar por completo esta materia en un capítulo, aclararemos algunos temas esenciales de la especificación y evaluación de modelos.

## 13.1 Criterios de selección del modelo

De acuerdo con Hendry y Richard, la elección de un modelo para el análisis empírico debe satisfacer los siguientes criterios:[4]

1. *Ser adecuado para los datos;* es decir, las predicciones basadas en el modelo deben ser lógicamente posibles.

2. *Ser consistente con la teoría;* es decir, debe tener un sentido económico pertinente. Por ejemplo, si es válida la **hipótesis del ingreso permanente** de Milton Friedman, se espera que el valor del intercepto en la regresión del consumo permanente sobre el ingreso permanente sea igual a cero.

3. *Tener regresoras exógenas débiles;* es decir, las variables explicativas, o regresoras, no deben estar correlacionadas con el término de error. Puede añadirse que, en algunas situaciones, las regresoras exógenas tal vez sean **estrictamente exógenas**. Una variable estrictamente exógena es independiente de los valores actuales, futuros y pasados del término de error.

4. *Mostrar constancia en los parámetros;* es decir, los valores de los parámetros deben ser estables. De otra forma el pronóstico se dificultará. Como explica Friedman: "La única prueba relevante de la validez de un[a] [modelo] hipótesis es la comparación de sus predicciones con la experiencia".[5] Ante la ausencia de la constancia en los parámetros, tales predicciones no serán confiables.

5. *Exhibir coherencia en los datos;* es decir, los residuos estimados a partir del modelo deben ser puramente aleatorios (técnicamente, ruido blanco). En otras palabras, si el modelo de regresión es adecuado, los residuos obtenidos de este modelo deben ser de ruido blanco. Si no es el caso, existe un error de especificación en el modelo. En breve exploraremos la naturaleza del (los) error(es) de especificación.

6. *Ser inclusivo;* es decir, el modelo debe abarcar o incluir todos los modelos contendientes, en el sentido de que debe poder explicar sus resultados. En resumen, otros modelos no pueden ser mejores que el elegido.

Una cosa es poner en una lista los criterios de un "buen" modelo y otra muy distinta desarrollarlos en realidad, pues en la práctica es muy probable que se cometan diversos errores de especificación en los modelos, que analizaremos en la siguiente sección.

## 13.2 Tipos de errores de especificación

Suponga que con base en los criterios recién enumerados llegamos a un modelo que aceptamos como bueno. En concreto, este modelo es

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i} \tag{13.2.1}$$

donde $Y =$ costo total de producción y $X =$ producción. La ecuación (13.2.1) es un ejemplo de la función cúbica del costo total frecuente en los libros de texto.

[4] D.F. Hendry y J.F. Richard, "The Econometric Analysis of Economic Time Series", *International Statistical Review,* vol. 51, 1983, pp. 3-33.

[5] Milton Friedman, "The Methodology of Positive Economics", en *Essays in Positive Economics,* University of Chicago Press, Chicago, 1953, p. 7.

Pero suponga que, por alguna razón (por ejemplo, por pereza de graficar el diagrama de dispersión), un investigador decide utilizar el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \qquad \textbf{(13.2.2)}$$

Observe que cambiamos la notación para distinguir este modelo del modelo verdadero.

Como se supone que (13.2.1) es verdadero, la adopción de (13.2.2) constituiría un error de especificación, que consiste en la **omisión de una variable relevante** ($X_i^3$). Por consiguiente, el término de error $u_{2i}$ en (13.2.2) es de hecho

$$u_{2i} = u_{1i} + \beta_4 X_i^3 \qquad \textbf{(13.2.3)}$$

Pronto veremos la importancia de esta relación.

Ahora suponga que otro investigador utiliza el siguiente modelo:

$$Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + \lambda_4 X_i^3 + \lambda_5 X_i^4 + u_{3i} \qquad \textbf{(13.2.4)}$$

Si (13.2.1) es el "verdadero" (13.2.4), también constituye un error de especificación que consiste en **incluir una variable innecesaria o irrelevante** en el sentido de que el modelo verdadero supone que $\lambda_5$ es cero. El nuevo término de error es de hecho

$$\begin{aligned} u_{3i} &= u_{1i} - \lambda_5 X_i^4 \\ &= u_{1i} \qquad \text{pues } \lambda_5 = 0 \text{ en el modelo verdadero} \qquad (\text{¿Por qué?}) \end{aligned} \qquad \textbf{(13.2.5)}$$

Suponga ahora que otro investigador postula el siguiente modelo:

$$\ln Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + u_{4i} \qquad \textbf{(13.2.6)}$$

En relación con el modelo verdadero (13.2.6), también presenta un sesgo de especificación, en este caso originado por una **forma funcional incorrecta:** en (13.2.1) $Y$ aparece linealmente, mientras que en (13.2.6) aparece en forma log-lineal.

Por último, considere que un investigador utiliza el siguiente modelo:

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4^* X_i^{*3} + u_i^* \qquad \textbf{(13.2.7)}$$

donde $Y_i^* = Y_i + \varepsilon_i$ y $X_i^* = X_i + w_i$, con $\varepsilon_i$ y $w_i$ como errores de medición. Lo que plantea (13.2.7) es que, en lugar de los verdaderos $Y_i$ y $X_i$, se utilizan sus valores sustitutos, $Y_i^*$ y $X_i^*$, los cuales pueden contener errores de medición. Por consiguiente, en (13.2.7) hay un **sesgo por errores de medición**. En el trabajo aplicado, la información está plagada de errores de aproximación, errores de cobertura incompleta o tan sólo errores de omisión de algunas observaciones. En las ciencias sociales a menudo se depende de datos secundarios y no hay forma de conocer los tipos de errores, si existen, cometidos por la agencia recolectora de datos primarios.

Otro tipo de error de especificación se relaciona con la forma como el error estocástico $u_i$ (o $u_t$) entra en el modelo de regresión. Considere por ejemplo el siguiente modelo de regresión bivariado sin término de intercepto:

$$Y_i = \beta X_i u_i \qquad \textbf{(13.2.8)}$$

donde el término de error estocástico entra de forma multiplicativa y tiene la propiedad de que ln $u_i$ satisface los supuestos del MCRL, en comparación con el siguiente modelo,

$$Y_i = \alpha X_i + u_i \qquad \textbf{(13.2.9)}$$

donde el término de error entra en forma aditiva. Aunque las variables son las mismas en ambos modelos, se denotó el coeficiente de la pendiente en (13.2.8) con $\beta$, y el coeficiente de la

pendiente en (13.2.9), con $\alpha$. Ahora, si (13.2.8) es el modelo "correcto" o "verdadero", ¿la $\alpha$ estimada proporciona una estimación insesgada de la verdadera $\beta$? Es decir, ¿será $E(\hat{\alpha}) = \beta$? Si no es el caso, la especificación estocástica inadecuada del término de error constituirá otra fuente de errores de especificación.

Un error de especificación que a veces se pasa por alto es la **interacción entre las regresoras**, es decir, el **efecto multiplicativo de una o más regresoras sobre la variable regresada**. Para ilustrar, considere la siguiente función simplificada de salarios:

$$\begin{aligned} \ln W_i = \beta_1 + \beta_2 \, \text{Escolaridad}_i + \beta_3 \, \text{Sexo}_i \\ + \beta_4 \, (\text{Escolaridad}) \, (\text{Sexo}) + u \end{aligned} \qquad (13.2.10)$$

En este modelo, el cambio en los salarios relativos respecto de la educación depende no sólo de la escolaridad sino también del sexo ($\frac{\partial \ln W}{\partial \text{Escolaridad}} = \beta_2 + \beta_4 \text{Sexo}$). Asimismo, el cambio en los salarios relativos respecto del sexo depende no sólo del sexo, sino también del nivel de escolaridad.

Para resumir, al formular un modelo empírico, es probable que se cometan uno o más de los siguientes errores de especificación:

1. Omisión de una variable relevante.
2. Inclusión de una variable innecesaria.
3. Adopción de la forma funcional incorrecta.
4. Errores de medición.
5. Especificación incorrecta del término de error estocástico.
6. Suposición de que el término de error está normalmente distribuido.

Antes de examinar con detalle esos errores de especificación, vale la pena distinguir entre **errores de especificación del modelo** y **errores de especificación incorrecta del modelo**. Los primeros cuatro tipos de error son en esencia errores de especificación del modelo, pues lo que se tiene en mente es un modelo "verdadero", sin embargo, no estimamos el modelo correcto. En los errores de especificación incorrecta del modelo, para empezar, ni siquiera sabemos cuál es el verdadero modelo. En este contexto viene a la mente la controversia entre los keynesianos y los monetaristas. Estos últimos dan preferencia al dinero cuando explican los cambios en el PIB, en tanto que los keynesianos destacan el papel del gasto gubernamental para justificar las variaciones del PIB. Así que podemos decir que hay dos modelos rivales.

En lo que resta del capítulo veremos primero los errores de especificación de modelos y luego los errores de la mala especificación de modelos.

## 13.3 Consecuencias de los errores de especificación del modelo

Independientemente de las fuentes de los errores de especificación, ¿cuáles son las consecuencias? Para no complicar este análisis, responderemos en el contexto del modelo con tres variables y consideraremos en detalle dos tipos de errores de especificación ya analizados, a saber: 1) **subajuste de un modelo**, es decir, la omisión de variables relevantes, y 2) **sobreajuste de un modelo**, es decir, la inclusión de variables innecesarias. Por supuesto, los resultados se pueden generalizar al caso de más de dos regresoras, pero mediante manejo algebraico tedioso;[6] una vez que se tienen casos de más de tres variables, el álgebra matricial se convierte en necesidad.

[6] Véase, sin embargo, el ejercicio 13.32.

## Omisión de una variable relevante (subajuste de un modelo)

Suponga que el verdadero modelo es

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{13.3.1}$$

pero, por alguna razón ajustamos el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i \tag{13.3.2}$$

Las consecuencias de omitir $X_3$ son las siguientes:

1. Si la variable excluida, u omitida, $X_3$ está correlacionada con la variable incluida $X_2$, es decir, $r_{23}$, el coeficiente de correlación entre las dos variables es *diferente de cero*, $\hat{\alpha}_1$ y $\hat{\alpha}_2$ son *sesgados e inconsistentes*. Es decir, $E(\hat{\alpha}_1)$ no es igual a $\beta_1$ y $E(\hat{\alpha}_2)$ no es igual a $\beta_2$, y el sesgo no desaparece conforme aumenta el tamaño de la muestra.

2. Aunque $X_2$ y $X_3$ no estén correlacionados, $\hat{\alpha}_1$ es sesgado, pese a que $\hat{\alpha}_2$ sea ahora insesgado.

3. La varianza de la perturbación $\sigma^2$ está incorrectamente estimada.

4. La varianza medida convencionalmente de $\hat{\alpha}_2\ (= \sigma^2/\sum x_{2i}^2)$ es un estimador *sesgado* de la varianza del verdadero estimador $\hat{\beta}_2$.

5. En consecuencia, es probable que el intervalo de confianza usual y los procedimientos de pruebas de hipótesis conduzcan a conclusiones equivocadas sobre la significancia estadística de los parámetros estimados.

6. Otra consecuencia es que los pronósticos basados en el modelo incorrecto y los intervalos (de confianza) del pronóstico no son confiables.

Aunque las pruebas de cada una de las afirmaciones anteriores escapan por mucho al tema,[7] en el apéndice 13A, sección 13A.1, se demuestra que

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{32} \tag{13.3.3}$$

donde $b_{32}$ es la pendiente en la regresión de la variable excluida $X_3$ sobre la variable incluida $X_2$ $(b_{32} = \sum x_{3i}x_{2i}/\sum x_{2i}^2)$. Como se ve en (13.3.3), $\hat{\alpha}_2$ está sesgada, a menos que $\beta_3$ o $b_{32}$, o ambas, sean cero. Eliminamos $\beta_3$ haciéndola cero porque en ese caso, para empezar, ni siquiera hay error de especificación. El coeficiente $b_{32}$ será cero si $X_2$ y $X_3$ no están correlacionadas, lo cual es poco probable en la mayoría de los datos económicos.

Sin embargo, por lo general, la amplitud del sesgo depende del *término del sesgo*, $\beta_3 b_{32}$. Si, por ejemplo, $\beta_3$ es positiva (es decir, $X_3$ tiene un efecto positivo sobre $Y$) y $b_{32}$ es positiva (es decir, $X_2$ y $X_3$ están positivamente correlacionadas), $\hat{\alpha}_2$, en promedio, sobreestimará a la verdadera $\beta_2$ (es decir, al sesgo positivo). Pero este resultado no debe sorprender, pues $X_2$ representa no solamente su *efecto directo* sobre $Y$ sino también su *efecto indirecto* (a través de $X_3$) sobre $Y$. En resumen, $X_2$ obtiene relevancia por la influencia que debe atribuirse a $X_3$, sin permitir que esta última muestre su efecto explícitamente porque no se le "permite" ingresar al modelo. Como ejemplo concreto, considere el que analizamos en el capítulo 7 (ejemplo 7.1).

[7] Para un tratamiento algebraico, véase Jan Kmenta, *Elements of Econometrics,* Macmillan, Nueva York, 1971, pp. 391-399. Quienes conozcan el álgebra matricial pueden consultar J. Johnston, *Econometric Methods,* 4a. ed., McGraw-Hill, Nueva York, 1997, pp. 119-122.

**EJEMPLO 13.1**
*Ejemplo ilustrativo: De nuevo la mortalidad infantil*

Al hacer la regresión de la mortalidad infantil (MI) sobre el PIB *per cápita* (PIBPC) y sobre la tasa de alfabetización de las mujeres (TAM) obtuvimos los resultados de la ecuación (7.6.2), con los valores parciales de los coeficientes de pendiente de las dos variables −0.0056 y −2.2316, respectivamente. Pero si ahora eliminamos la variable (TAM) obtenemos los resultados de la ecuación (7.7.2). Si consideramos que (7.6.2) es el modelo correcto, entonces (7.7.2) es un modelo mal especificado, pues omite la variable relevante TAM. Ahora podemos observar que, en el modelo correcto, el coeficiente de la variable PIBPC fue −0.0056, en tanto que en el modelo "incorrecto" (7.7.2) es ahora de −0.0114.

En términos absolutos, ahora la variable PIBPC tiene un mayor impacto sobre la MI en comparación con el verdadero modelo. Pero si hacemos la regresión de TAM sobre PIBPC (la regresión de la variable excluida sobre la incluida), el coeficiente de pendiente en la regresión [$b_{32}$ en términos de la ecuación (13.3.3)] es 0.00256.[8] Lo anterior indica que conforme PIBPC aumenta una unidad, en promedio, TAM se incrementa 0.00256 unidades. Pero si TAM aumenta esas unidades, su efecto en MI será $(-2.2316)(0.00256) = \hat{\beta}_3 b_{32} = -0.00543$.

Por tanto, de (13.3.3) tenemos al final $(\hat{\beta}_2 + \hat{\beta}_3 b_{32}) = [-0.0056 + (-2.2316)(0.00256)] \approx -0.0111$, que es casi el valor del coeficiente PIBPC, obtenido en el modelo incorrecto (7.7.2).[9] Como ilustra este ejemplo, el verdadero efecto del PIBPC sobre la MI es mucho menor (−0.0056) de lo que indica el modelo incorrecto (7.7.2), a saber, (−0.0114).

Examinemos ahora las varianzas de $\hat{\alpha}_2$ y $\hat{\beta}_2$

$$\operatorname{var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \tag{13.3.4}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} = \frac{\sigma^2}{\sum x_{2i}^2}\text{FIV} \tag{13.3.5}$$

donde FIV (una medida de colinealidad) es el factor inflacionario de la varianza $[= 1/(1 - r_{23}^2)]$ analizada en el capítulo 10 y $r_{23}$ es el coeficiente de correlación entre las variables $X_2$ y $X_3$; las ecuaciones (13.3.4) y (13.3.5) resultan familiares pues las vimos en los capítulos 3 y 7.

Como las fórmulas (13.3.4) y (13.3.5) no son iguales, en general la var($\hat{\alpha}_2$) será diferente de la var($\hat{\beta}_2$). Pero sabemos que var($\hat{\beta}_2$) es insesgada (¿por qué?). Por tanto, var($\hat{\alpha}_2$) es sesgada, de modo que justifica el enunciado del punto 4 anterior. Como $0 < r_{23}^2 < 1$, parece que en el presente caso var($\hat{\alpha}_2$) < ($\hat{\beta}_2$). Ahora surge un dilema: aunque $\hat{\alpha}_2$ sea sesgada, su varianza es más pequeña que la varianza del estimador $\hat{\beta}_2$ insesgado (por supuesto, desechamos el caso en el que $r_{23} = 0$, pues en la práctica existe cierto grado de correlación entre las regresoras). Por consiguiente, se trata de una situación en que hay que sacrificar una cosa por otra.[10]

Pero aún no termina esto, pues la $\sigma^2$ estimada del modelo (13.3.2) y la estimada del modelo verdadero (13.3.1) no son iguales, ya que la SCR de los dos modelos, así como sus grados de libertad (gl) son distintos. Recordará que obtuvimos un estimado de $\sigma^2$ como $\hat{\sigma}^2 = \text{SCR/gl}$, lo cual depende del número de regresoras incluidas en el modelo, así como de los gl (= $n$,

[8] Los resultados de la regresión son:

$$\widehat{\text{TAM}} = 47.5971 + 0.00256\text{PIBPC}$$
$$\text{ee} = (3.5553) \quad (0.0011) \qquad r^2 = 0.0721$$

[9] Observe que en el modelo verdadero $\hat{\beta}_2$ y $\hat{\beta}_3$ son estimaciones insesgadas de sus valores verdaderos o reales.

[10] A fin de superar el dilema entre el sesgo y la eficiencia, se puede elegir reducir el error cuadrático medio (ECM), pues se relaciona con el sesgo y la eficiencia. Sobre el ECM, véase el apéndice estadístico (**apéndice A**). También consulte el ejercicio 13.6.

número de parámetros estimados). Ahora bien, si añadimos variables al modelo, por lo general la SCR decrece (recuerde que mientras más variables se añadan al modelo, más se incrementa $R^2$), pero los grados de libertad también disminuyen porque se estiman más parámetros. El resultado global depende de que la SCR decrezca lo suficiente para compensar la pérdida de grados de libertad debido a la incorporación de regresoras. Es muy probable que si una regresora tiene un gran impacto sobre la regresada —por ejemplo, puede reducir la SCR en mayor medida de lo que significa la pérdida de grados de libertad como resultado de incorporarse al modelo—, la inclusión de tales variables no sólo reduce el sesgo, sino que también aumenta la precisión (es decir, disminuye los errores estándar) de los estimadores.

Por otra parte, si las variables relevantes sólo tienen un efecto marginal en la regresada, y si están muy correlacionadas (es decir, el FIV es mayor), se puede reducir el sesgo en los coeficientes de las variables ya incluidas en el modelo, pero aumentarían sus errores estándar (es decir, se harían menos eficientes). De hecho, la disyuntiva entre mejor precisión o menos sesgo, en esta situación, puede ser sustancial. Como se desprende del análisis, la decisión depende de la importancia relativa de las diversas regresoras.

Para concluir, consideremos ahora el caso especial en donde $r_{23} = 0$, es decir, $X_2$ y $X_3$ no están correlacionadas. En este caso, $b_{32}$ es cero (¿por qué?). Por consiguiente, se ve, de (13.3.3), que $\hat{\alpha}_2$ es ahora insesgada.[11] También, de (13.3.4) y (13.3.5) parece que las varianzas de $\hat{\alpha}_2$ y $\hat{\beta}_2$ son las mismas. ¿No hay perjuicio entonces en eliminar la variable $X_3$ del modelo aunque pueda ser relevante en teoría? La respuesta suele ser negativa, pues, en este caso, como ya indicamos, la var($\hat{\alpha}_2$) estimada de (13.3.4) es aún sesgada y, por consiguiente, es probable que los procedimientos de pruebas de hipótesis continúen siendo dudosos.[12] Además, en la mayoría de investigaciones económicas es probable que $X_2$ y $X_3$ estén correlacionadas, lo que crea los problemas mencionados. **El punto es muy claro: una vez formulado el modelo con base en la teoría pertinente, no se aconseja eliminar una variable de dicho modelo.**

## Inclusión de una variable irrelevante (sobreajuste de un modelo)

Ahora supongamos que

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \tag{13.3.6}$$

es verdadero, pero especificamos el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i \tag{13.3.7}$$

y cometemos así el error de especificación al incluir una variable innecesaria en el modelo.

Las consecuencias de este error de especificación son las siguientes:

1. Todos los estimadores de MCO de los parámetros del modelo "incorrecto" son insesgados y consistentes, es decir, $E(\alpha_1) = \beta_1$, $E(\hat{\alpha}_2) = \beta_2$ y $E(\hat{\alpha}_3) = \beta_3 = 0$.

2. La varianza del error $\sigma^2$ está correctamente estimada.

3. Los procedimientos usuales de intervalos de confianza y de pruebas de hipótesis conservan su validez.

4. Sin embargo, las $\alpha$ estimadas por lo general serán ineficientes, es decir, sus varianzas generalmente serán más grandes que las de las $\hat{\beta}$ del verdadero modelo. Las pruebas de algunas de estas afirmaciones se encuentran en el apéndice 13A, sección 13A.2. El punto de interés aquí es la ineficiencia relativa de las $\hat{\alpha}$. Esto se demuestra fácilmente.

---

[11] Observe, sin embargo, que $\hat{\alpha}_1$ es aún sesgado, lo cual se ve intuitivamente de la siguiente manera: sabemos que $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$, mientras que $\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2 \bar{X}_2$, y aunque $\hat{\alpha}_2 = \hat{\beta}_2$, los dos estimadores no serán iguales.

[12] Para mayores detalles, véase Adrian C. Darnell, *A Dictionary of Econometrics,* Edward Elgar Publisher, 1994, pp. 371-372.

De la fórmula usual de MCO sabemos que

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \tag{13.3.8}$$

y

$$\operatorname{var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \tag{13.3.9}$$

Por consiguiente,

$$\frac{\operatorname{var}(\hat{\alpha}_2)}{\operatorname{var}(\hat{\beta}_2)} = \frac{1}{1 - r_{23}^2} \tag{13.3.10}$$

Como $0 \le r_{23}^2 \le 1$, se cumple que $\operatorname{var}(\hat{\alpha}_2) \ge \operatorname{var}(\hat{\beta}_2)$, es decir, la varianza de $\hat{\alpha}_2$ suele ser más grande que la varianza de $\hat{\beta}_2$, aunque, en promedio, $\hat{\alpha}_2 = \beta_2$ [es decir, $E(\hat{\alpha}_2) = \beta_2$].

La implicación de este hallazgo es que la inclusión de la variable innecesaria $X_3$ hace que la varianza de $\hat{\alpha}_2$ sea más grande de lo necesario, con lo cual $\hat{\alpha}_2$ se hace menos precisa. Esto también es cierto con $\hat{\alpha}_1$.

Observe la **asimetría** en los dos tipos de sesgos de especificación que consideramos. Si excluimos una variable relevante, los coeficientes de las variables consideradas en el modelo son por lo general sesgados e inconsistentes, la varianza del error es incorrectamente estimada y se invalidan los procedimientos usuales de pruebas de hipótesis. Por otra parte, la inclusión de una variable irrelevante en el modelo proporciona estimaciones insesgadas y consistentes de los coeficientes en el modelo verdadero, la varianza del error es correctamente estimada y los métodos convencionales de pruebas de hipótesis son aún válidos; la única penalización por la inclusión de la variable superflua es que las varianzas estimadas de los coeficientes son mayores y, como resultado, las inferencias probabilísticas sobre los parámetros son menos precisas. Una conclusión no deseada aquí sería que es mejor incluir variables irrelevantes que omitir variables relevantes. Pero esta filosofía no es estricta, pues incluir variables innecesarias genera una pérdida de eficiencia de los estimadores y puede provocar también el problema de multicolinealidad (¿por qué?), para no mencionar la pérdida de grados de libertad. Por consiguiente,

> En general, el mejor enfoque es incluir sólo las variables explicativas que, en teoría, influyan *directamente* en la variable dependiente y no se hayan tomado en cuenta en otras variables incluidas.[13]

## 13.4 Pruebas de errores de especificación

Conocer las consecuencias de los errores de especificación es una cosa, pero averiguar si se cometieron tales errores es otra muy diferente, pues en la especificación no se espera deliberadamente cometer estos errores. Con mucha frecuencia, los sesgos de especificación surgen en forma inadvertida, quizá por la incapacidad de formular el modelo en la forma más precisa posible debido a que la teoría es débil o a que no se tiene la clase de información adecuada para probar el modelo. Como observa Davidson: "Debido a la naturaleza no experimental de la economía, nunca estamos seguros de la forma en que se generaron los datos observados. En economía, resulta que la prueba de cualquier hipótesis siempre depende de supuestos adicionales necesarios para especificar un modelo razonablemente ambicioso, los cuales pueden o no estar justificados".[14]

---

[13] Michael D. Intriligator, *Econometric Models, Techniques and Applications,* Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1978, p. 189. Recuerde el principio de la navaja de Occam.

[14] James Davidson, *Econometric Theory,* Blackwell Publishers, Oxford, Inglaterra, 2000, p. 153.

La pregunta práctica no es por qué se cometen tales errores, pues por lo general los hay, sino cómo detectarlos. Una vez que se encuentran errores de especificación, con frecuencia los remedios surgen por sí mismos. Si, por ejemplo, puede demostrarse que una variable se omitió inapropiadamente de un modelo, el remedio obvio es incluirla en el análisis, suponiendo que, desde luego, se tenga información disponible sobre ella.

En esta sección analizamos algunas pruebas para detectar errores de especificación.

## Detección de variables innecesarias (sobreajuste de un modelo)

Suponga que desarrollamos un modelo de $k$ variables para explicar un fenómeno:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{13.4.1}$$

Sin embargo, no tenemos una certeza total de que, por ejemplo, la variable $X_k$ debe estar en el modelo. Una forma sencilla de averiguarlo es probar la significancia del $\beta_k$ estimado mediante la prueba $t$ usual: $t = \hat{\beta}_k/\text{ee}\,(\hat{\beta}_k)$. Pero suponga que no hay seguridad de que $X_3$ y $X_4$ pertenezcan en realidad al modelo. Esto se consigue fácilmente mediante la prueba $F$ estudiada en el capítulo 8. Así, la detección de una o más variables irrelevantes no es difícil.

Pero es muy importante recordar que, con estas pruebas de significancia, tenemos en mente un modelo específico. Aceptamos ese modelo como **hipótesis mantenida** o "verdad", sin importar lo tentativa que pueda ser. Así, con ese modelo, mediante las pruebas usuales $t$ o $F$ podemos averiguar la relevancia verdadera de una o más regresoras. Pero observe con cuidado que con las pruebas $t$ y $F$ no podemos construir un modelo en forma iterativa, es decir, no podemos decir que al principio $Y$ está relacionada con $X_2$ sólo porque $\hat{\beta}_2$ es estadísticamente significativa, ampliar luego el modelo para incluir $X_3$ y luego conservar esa variable en el modelo si $\hat{\beta}_3$ resulta ser estadísticamente significativo, y así sucesivamente. Esta estrategia de elaborar un modelo se llama **método ascendente** (se empieza con un modelo más pequeño y se amplía conforme se prosigue) o, un término más descriptivo: **minería de datos** (otros nombres son **regresión al tanteo, extracción de datos, sondeo de datos** y **procesamiento masivo de datos numéricos**).

El objetivo principal de la minería de datos es desarrollar el "mejor" modelo después de varias pruebas de diagnóstico, de manera que el modelo final resulte "bueno" en el sentido de que todos los coeficientes estimados tengan los signos "correctos", sean estadísticamente significativos de acuerdo con las pruebas $t$ y $F$, el valor $R^2$ resulte razonablemente alto y el $d$ de Durbin-Watson tenga un valor aceptable (alrededor de 2), etc. Los puristas menosprecian la práctica de la minería de datos. En palabras de William Pool, ". . . siempre resulta arriesgado hacer de una regularidad empírica el fundamento, en vez de tomar como base una implicación de la teoría económica".[15] En seguida daremos una razón para "condenar" la minería de datos.

### *Nivel de significancia nominal frente a nivel de significancia verdadero en presencia de minería de datos*

Un peligro de la minería de datos al cual se enfrenta el investigador desprevenido es que los niveles convencionales de significancia ($\alpha$) como 1, 5 o 10% *no son los verdaderos niveles de significancia*. Lovell sugirió que, si hay $c$ candidatas regresoras de las cuales $k$ son finalmente seleccionadas ($k \leq c$) con base en la minería de datos, el verdadero nivel de significancia ($\alpha^*$) se relaciona con el nivel de significancia nominal ($\alpha$) de la siguiente manera:[16]

$$\alpha^* = 1 - (1 - \alpha)^{c/k} \tag{13.4.2}$$

[15] William Pool, "Is Inflation Too Low?", *Cato Journal,* vol. 18, núm. 3, invierno de 1999, p. 456.

[16] M. Lovell, "Data Mining", *Review of Economics and Statistics,* vol. 65, 1983, pp. 1-12.

o aproximadamente como

$$\alpha^* \approx (c/k)\alpha \tag{13.4.3}$$

Por ejemplo, si $c = 15$, $k = 5$ y $\alpha = 5\%$, mediante (13.4.3), el verdadero nivel de significancia es $(15/5)(5) = 15\%$. Por consiguiente, si un investigador extrae datos, selecciona 5 de 15 regresoras y sólo informa los resultados en el nivel de significancia de 5% nominal, y declara que estos resultados son estadísticamente significativos, esta conclusión se debe tomar con gran reserva; hasta donde sabemos, el (verdadero) nivel de significancia es en realidad 15%. Debe observarse que si $c = k$, es decir, si no se ha hecho minería de datos, los niveles de significancia verdadero y nominal son iguales. Por supuesto, en la práctica la mayoría de los investigadores sólo informa los resultados de su regresión "final" sin reconocer que llegaron a los resultados tras una considerable minería de datos, o **preprueba**.[17]

Aparte de algunas desventajas evidentes, cada vez se reconoce más, sobre todo los econometristas aplicados, que el método purista (es decir, el que no realiza minería de datos) para elaborar modelos no es defendible. Como expresa Zaman:

> Por desgracia, la experiencia con los conjuntos de datos reales muestra que tal enfoque [el purista] no es factible ni deseable. No es factible porque es una teoría económica extraña que conduce a un modelo único. No es deseable porque un aspecto crucial del aprendizaje mediante los datos es conocer los tipos de modelos que los datos apoyan o rechazan. Aunque, por una extraña suerte, el modelo inicial mostrase un buen ajuste, con frecuencia resultará importante explorar y conocer las clases de modelos con que los datos concuerdan o no.[18]

Kerry Patterson expresa un punto de vista similar:

> Este enfoque [el de minería de datos] indica que la teoría económica y la especificación empírica interactúan en vez de mantenerse en compartimentos separados.[19]

En lugar de enredarse con la disyuntiva entre la minería de datos y el enfoque purista para la elaboración de modelos, uno puede inclinarse por la posición de Peter Kennedy:

> [la especificación de modelos] necesitaría ser una combinación bien pensada de teoría y datos; asimismo, los procedimientos de prueba para la búsqueda de especificaciones deben estar diseñados para reducir los costos de la minería de datos. Ejemplos de dichos procedimientos son: dejar de lado datos para las pruebas de predicción fuera de las muestras, ajustar los niveles de significancia [a la Lovell] y evitar criterios cuestionables, como maximizar $R^2$.[20]

Si consideramos la minería de datos desde una perspectiva más amplia, como un proceso de descubrimiento de regularidades empíricas que sugiriese errores y/u omisiones en los modelos teóricos (existentes), quizá desempeñara un papel muy útil. Para citar de nueva cuenta a Kennedy, "El arte del econometrista aplicado consiste en permitir que la teoría se deje conducir por los datos y al mismo tiempo evite los enormes daños que implica la minería de datos".[21]

---

[17] Hay un análisis detallado de lo que pueden provocar las prepruebas y los sesgos en T.D. Wallace, "Pretest Estimation in Regression: A Survey", *American Journal of Agricultural Economics,* vol. 59, 1977, pp. 431-443.

[18] Asad Zaman, *Statistical Foundations for Econometric Techniques,* Academic Press, Nueva York, 1996, p. 226.

[19] Kerry Patterson, *An Introduction to Applied Econometrics,* St. Martin's Press, Nueva York, 2000, p. 10.

[20] Peter Kennedy, "Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics", manuscrito inédito.

[21] Kennedy, *op. cit.,* p. 13.

## Pruebas para variables omitidas y forma funcional incorrecta

En la práctica, nunca estamos seguros de que el modelo adoptado para pruebas empíricas represente "la verdad, toda la verdad y nada más que la verdad". Con base en la teoría o en la introspección y en el trabajo empírico previo, desarrollamos un modelo que se cree recoge la esencia del tema en estudio. Luego sometemos el modelo a una prueba empírica. Después de obtener los resultados, iniciamos el *post mortem*, teniendo en mente los criterios ya estudiados de un buen modelo. Es en esta etapa cuando nos enteramos si el modelo seleccionado es adecuado. Al determinar la bondad de ajuste del modelo se observan algunas características generales de los resultados, como el valor $\bar{R}^2$, las razones $t$ estimadas, los signos de los coeficientes estimados en relación con sus expectativas previas, el estadístico de Durbin-Watson, etc. Si estos diagnósticos son razonablemente buenos, podemos afirmar que el modelo seleccionado es una buena representación de la realidad. Con el mismo procedimiento, si los resultados no parecen estimulantes porque el valor de $\bar{R}^2$ es muy bajo o porque muy pocos coeficientes son estadísticamente significativos o tienen los signos correctos, o porque el $d$ de Durbin-Watson es muy bajo, entonces puede empezar a preocupar la bondad del ajuste del modelo y podemos empezar a buscar remedios: tal vez omitimos una variable importante, utilizamos la forma funcional equivocada o no realizamos la primera diferenciación de la serie de tiempo (para eliminar la correlación serial), y así sucesivamente. Para determinar si la incompetencia del modelo se debe a uno o más de estos problemas están algunos de los siguientes métodos.

### *Examen de los residuos*

Como mencionamos en el capítulo 12, el examen de los residuos es un buen diagnóstico visual para detectar la autocorrelación o la heteroscedasticidad. Pero estos residuos también se examinan, en especial en información de corte transversal, para detectar errores de especificación en los modelos, como la omisión de una variable importante o la definición de una forma funcional incorrecta. Si en realidad existen tales errores, una gráfica de los residuos permite apreciar patrones distinguibles.

Para ilustrar lo anterior, reconsidere la función cúbica del costo total de producción analizada en el capítulo 7. Suponga que la verdadera función del costo total se describe de la siguiente manera, donde $Y =$ costo total y $X =$ producción:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i \tag{13.4.4}$$

pero un investigador ajusta la siguiente función cuadrática:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \tag{13.4.5}$$

y otro investigador ajusta la siguiente función lineal:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \tag{13.4.6}$$

Aunque sabemos que ambos investigadores cometieron errores de especificación, con fines pedagógicos veamos cómo se comportan los residuos estimados en los tres modelos. (La información costo-producción está en la tabla 7.4.) La figura 13.1 habla por sí misma: a medida que nos movemos de izquierda a derecha, es decir, a medida que nos acercamos a la verdad, no sólo los residuos son más pequeños (en valor absoluto) sino, asimismo, éstos no presentan los giros cíclicos pronunciados asociados con modelos mal especificados.

La utilidad de examinar la gráfica de residuos es entonces clara: si hay errores de especificación, los residuos presentan patrones distinguibles.

### *De nuevo, el estadístico d de Durbin-Watson*

Si examinamos el estadístico $d$ de Durbin-Watson que se calcula de manera habitual y aparece en la tabla 13.1, vemos que, para la función lineal de costos, el $d$ estimado es 0.716, lo cual indica que hay "correlación" positiva en los residuos estimados: para $n = 10$ y $k' = 1$, los valores $d$

**FIGURA 13.1**
Residuos $\hat{u}_i$ obtenidos de las funciones del costo total de tipo *a*) lineal, *b*) cuadrática y *c*) cúbica.

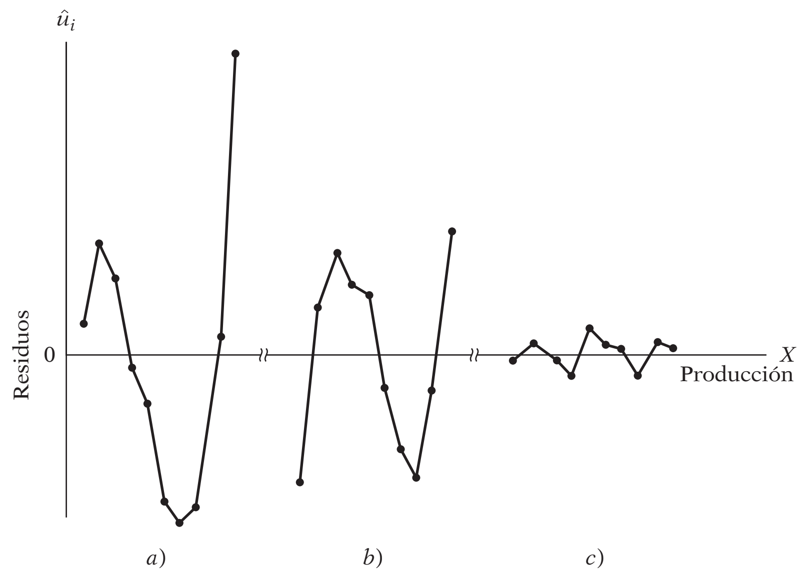



**TABLA 13.1**
**Residuos estimados de las funciones del costo total de tipo lineal, cuadrático y cúbico**

| Número de observación | $\hat{u}_i$, modelo lineal* | $\hat{u}_i$, modelo cuadrático† | $\hat{u}_i$, modelo cúbico** |
|---|---|---|---|
| 1 | 6.600 | −23.900 | −0.222 |
| 2 | 19.667 | 9.500 | 1.607 |
| 3 | 13.733 | 18.817 | −0.915 |
| 4 | −2.200 | 13.050 | −4.426 |
| 5 | −9.133 | 11.200 | 4.435 |
| 6 | −26.067 | −5.733 | 1.032 |
| 7 | −32.000 | −16.750 | 0.726 |
| 8 | −28.933 | −23.850 | −4.119 |
| 9 | 4.133 | −6.033 | 1.859 |
| 10 | 54.200 | 23.700 | 0.022 |

*$\hat{Y}_i = 166.467 + 19.933X_i$ $\quad R^2 = 0.8409$
(19.021) (3.066) $\quad \bar{R}^2 = 0.8210$
(8.752) (6.502) $\quad d = 0.716$

†$\hat{Y}_i = 222.383 - 8.0250X_i + 2.542X_i^2$ $\quad R^2 = 0.9284$
(23.488) (9.809) (0.869) $\quad \bar{R}^2 = 0.9079$
(9.468) (−0.818) (2.925) $\quad d = 1.038$

**$\hat{Y}_i = 141.767 + 63.478X_i - 12.962X_i^2 + 0.939X_i^3$ $\quad R^2 = 0.9983$
(6.375) (4.778) (0.9856) (0.0592) $\quad \bar{R}^2 = 0.9975$
(22.238) (13.285) (−13.151) (15.861) $\quad d = 2.70$

críticos a 5% son $d_L = 0.879$ y $d_U = 1.320$. De la misma manera, el valor *d* calculado para la función cuadrática de costos es 1.038, mientras que los valores críticos a 5% son $d_L = 0.697$ y $d_U = 1.641$, lo cual señala indecisión. Pero con la prueba *d* modificada (véase el capítulo 12) podemos decir que hay "correlación" positiva en los residuos, pues el *d* calculado es menor que $d_U$. Para la función cúbica de costo, la verdadera especificación, el valor *d* estimado no indica "correlación" positiva alguna en los residuos.[22]

La "correlación" positiva observada en los residuos cuando ajustamos el modelo lineal o cuadrático no es una medida de correlación serial (de primer orden) sino del error (o errores) de

[22] En el contexto presente, un valor *d* = 2 significa que no hay error de especificación. (¿Por qué?)

especificación (del modelo). La correlación observada tan sólo refleja que hay una o más variables pertenecientes al modelo incluidas en el término de error y necesitan desecharse de éste e introducirse, por derecho propio, como variables explicativas: si excluimos $X_i^3$ de la función de costos, entonces, como lo muestra (13.2.3), el término de error en el modelo mal especificado (13.2.2) es en realidad $(u_{1i} + \beta_4 X_i^3)$, el cual presenta un patrón sistemático (por ejemplo, de autocorrelación positiva) si en realidad $X_i^3$ afecta a $Y$ significativamente.

Para aplicar la prueba de Durbin-Watson para detectar error (o errores) de especificación de un modelo, procedemos de la siguiente manera:

1. A partir de un modelo supuesto, obtenga los residuos de MCO.

2. Si se cree que el modelo supuesto está mal especificado porque excluye una variable explicativa relevante, por ejemplo, $Z$, ordene los residuos obtenidos en el paso 1 de acuerdo con los valores crecientes de $Z$. *Nota:* La variable $Z$ puede ser una de las variables $X$ incluidas en el modelo supuesto o algún tipo de función de esa variable, como $X^2$ o $X^3$.

3. Calcule el estadístico $d$ a partir de los residuos así ordenados mediante la fórmula $d$ usual, a saber,

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2}$$

*Nota:* En este contexto, el subíndice $t$ es el índice de la observación que no necesariamente se refiere a una serie de tiempo.

4. Con base en las tablas de Durbin-Watson, si el valor $d$ estimado es significativo, se puede aceptar la hipótesis de mala especificación del modelo. Si es así, las medidas correctivas surgen naturalmente por sí mismas.

En el ejemplo de costos, la variable $Z(=X)$ (producción) ya fue ordenada.[23] Por consiguiente, no es preciso calcular otra vez el estadístico $d$. Como vimos, el estadístico $d$ para las funciones de costos lineal y cuadrática indica la presencia de errores de especificación. Los remedios son claros: introduzca los términos cuadrático y cúbico en la función lineal de costos y el término cúbico en la función cuadrática de costos. En resumen, efectúe la regresión del modelo cúbico de costos.

### *Prueba RESET de Ramsey*

Ramsey propuso una prueba general de errores de especificación conocida como RESET (prueba del error de especificación en regresión).[24] Aquí sólo ilustraremos la versión más sencilla de la prueba. Para establecer los conceptos, continuaremos con el ejemplo costo-producción y supondremos que la función de costos es lineal en la producción de la siguiente forma:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \tag{13.4.6}$$

donde $Y =$ costo total y $X =$ producción. Ahora, si graficamos los residuos $\hat{u}_i$ obtenidos de esta regresión frente a $\hat{Y}_i$, la estimación de $Y_i$ de este modelo, obtenemos la gráfica de la figura 13.2. Aunque $\sum \hat{u}_i$ y $\sum \hat{u}_i \hat{Y}_i$ necesariamente son cero (¿por qué?, véase el capítulo 3), los residuos en esta figura muestran un patrón en el cual su media cambia sistemáticamente con $\hat{Y}_i$. Esto indicaría que si introdujéramos $\hat{Y}_i$ en alguna forma como regresora(s) en (13.4.6), debería incrementar $R^2$. Y si el incremento en $R^2$ es estadísticamente significativo (con base en la prueba $F$ analizada en

[23] No importa si se ordena $\hat{u}_i$ de acuerdo con $X_i^2$ o con $X_i^3$, pues son funciones de $X_i$, la cual ya se ordenó.

[24] J.B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society,* serie B, vol. 31, 1969, pp. 350-371.

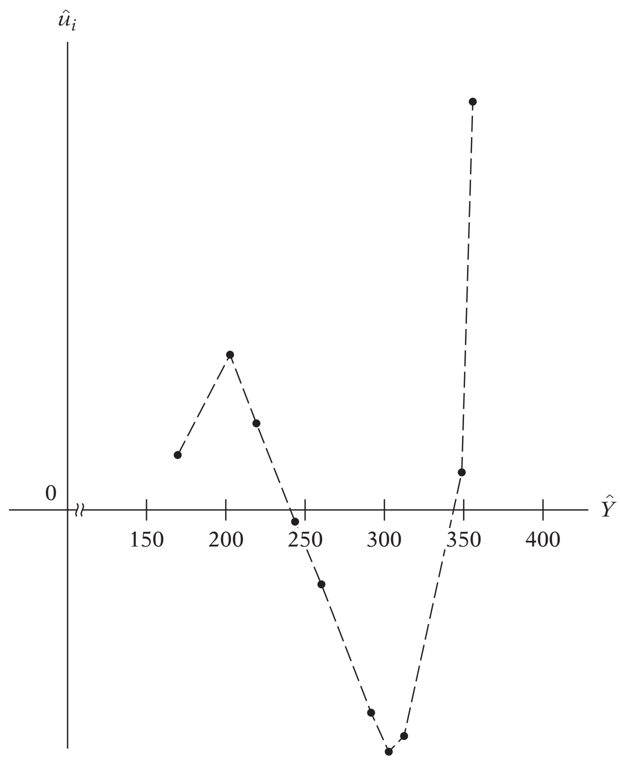



**FIGURA 13.2**
Residuos $\hat{u}_i$ y $Y$ estimados de la función lineal de costos: $Y_i = \lambda_1 + \lambda_2 X_i + u_i$.

el capítulo 8), esto sugeriría que la función lineal de costos (13.4.6) estaba mal especificada. Esta es la idea esencial de la prueba RESET. Los pasos de RESET son los siguientes:

1. A partir del modelo seleccionado, por ejemplo, el (13.4.6), obtenga $Y_i$ estimada, es decir, $\hat{Y}_i$.

2. Efectúe de nuevo la regresión (13.4.6) introduciendo $\hat{Y}_i$ en alguna forma, como una o varias regresoras adicionales. En la figura 13.2, observamos una relación curvilínea entre $\hat{u}_i$ y $\hat{Y}_i$, que indica que se pueden introducir $\hat{Y}_i^2$ y $\hat{Y}_i^3$ como regresoras adicionales. Así, efectuamos la regresión

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i \tag{13.4.7}$$

3. Sea $R^2$ obtenida de (13.4.7) $R^2_{\text{nueva}}$, y la obtenida de (13.4.6), $R^2_{\text{vieja}}$. Entonces utilizamos la prueba $F$ introducida ya en (8.4.18), a saber,

$$F = \frac{\left(R^2_{\text{nueva}} - R^2_{\text{vieja}}\right)/\text{número de regresoras nuevas}}{\left(1 - R^2_{\text{nueva}}\right)/(n - \text{número de parámetros en el nuevo modelo})} \tag{8.4.18}$$

para averiguar si el incremento en $R^2$, con (13.4.7), es estadísticamente significativo.

4. Si el valor $F$ calculado es significativo, por ejemplo, en el nivel de 5%, podemos aceptar la hipótesis de que el modelo (13.4.6) está mal especificado.

De regreso en el ejemplo ilustrativo, tenemos los siguientes resultados (los errores estándar están entre paréntesis):

$$\begin{aligned} \hat{Y}_i &= 166.467 \ + 19.933 X_i \\ &\quad (19.021) \quad (3.066) \qquad R^2 = 0.8409 \end{aligned} \tag{13.4.8}$$

$$\hat{Y}_i = 2\,140.7223 \;+ 476.6557X_i - \;0.09187\hat{Y}_i^2 + \;0.000119\hat{Y}_i^3$$
$$(132.0044) \quad (33.3951) \quad (0.00620) \quad (0.0000074) \qquad \textbf{(13.4.9)}$$
$$R^2 = 0.9983$$

*Nota:* $\hat{Y}_i^2$ y $\hat{Y}_i^3$ en (13.4.9) se obtienen de (13.4.8).

Ahora, al aplicar la prueba $F$, tenemos que

$$F = \frac{(0.9983 - 0.8409)/2}{(1 - 0.9983)/(10 - 4)} \qquad \textbf{(13.4.10)}$$
$$= 284.4035$$

El lector puede verificar fácilmente que este valor $F$ es muy significativo, lo cual indica que el modelo (13.4.8) está mal especificado. Por supuesto, llegamos a la misma conclusión con el examen visual de los residuos como también con el valor $d$ de Durbin-Watson. Debe añadirse que, en vista de que $\hat{Y}_i$ es estimada, se trata de una variable aleatoria y, por tanto, las pruebas de significancia habituales aplican si la muestra es razonablemente grande.

Una ventaja de RESET es que es fácil de aplicar, pues no requiere la especificación del modelo alterno. Sin embargo, ésta también es su desventaja, pues saber que el modelo está mal especificado no necesariamente ayuda a elegir una opción mejor.

Como apunta un autor:

> En la práctica, la prueba RESET puede no ser particularmente buena para detectar algguna alternativa específica para un modelo propuesto, y su utilidad radica en que sirve como indicador general de que algo está mal. Por esta razón, una prueba como RESET se describe en ocasiones como una prueba de especificación incorrecta en lugar de una prueba de especificación. Esta distinción es muy sutil, pero la idea básica es que una prueba de especificación examina algún aspecto particular de una ecuación dada, teniendo en mente hipótesis claras nula y alternativa. Una prueba de especificación incorrecta, por otra parte, puede detectar varias opciones e indica que algo está mal según la hipótesis nula, sin ofrecer necesariamente guía clara en cuanto a la hipótesis alterna apropiada.[25]

### *Prueba del multiplicador de Lagrange (ML) para agregar variables*

Ésta es una alternativa para la prueba RESET de Ramsey. Para explicar esta prueba, continuaremos con el ejemplo ilustrativo anterior.

Si comparamos la función lineal de costos (13.4.6) con la función cúbica de costos (13.4.4), la primera es una *versión restringida* de la última (recuerde el análisis de **mínimos cuadrados restringidos**, del capítulo 8). La regresión restringida (13.4.6) supone que los coeficientes de los términos de producción elevados al cuadrado y al cubo son iguales a cero. Para probar esto, la prueba ML se realiza de la siguiente manera:

1. Estime la regresión restringida (13.4.6) mediante MCO y obtenga los residuos, $\hat{u}_i$.
2. Si la regresión no restringida (13.4.4) resulta ser la verdadera regresión, los residuos obtenidos en (13.4.6) deben estar relacionados con los términos de la producción elevada al cuadrado y al cubo, es decir, $X_i^2$ y $X_i^3$.
3. Esto indica que se efectúe la regresión de los $\hat{u}_i$ obtenidos en el paso 1 sobre todas las regresoras (incluidas las de la regresión restringida), lo cual, en el presente caso, significa que

$$\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i \qquad \textbf{(13.4.11)}$$

donde $v$ es un término de error con las propiedades usuales.

[25] Jon Stewart y Len Gill, *Econometrics,* 2a. ed., Prentice-Hall Europe, 1998, p. 69.

4. Para un tamaño de muestra grande, Engle demostró que $n$ (el tamaño de la muestra) multiplicado por $R^2$ estimado en la regresión (auxiliar) (13.4.11) sigue una distribución ji cuadrada con gl iguales al número de restricciones impuestas por la regresión restringida, dos en el ejemplo presente, pues los términos $X_i^2$ y $X_i^3$ son eliminados del modelo.[26] Simbólicamente, escribimos

$$nR^2 \underset{\text{asin}}{\sim} \chi^2_{(\text{número de restricciones})} \tag{13.4.12}$$

donde asin significa asintóticamente, es decir, en muestras grandes.

5. Si el valor ji cuadrada obtenido de (13.4.12) excede el valor ji cuadrada crítico en el nivel de significancia seleccionado, rechazamos la regresión restringida. De lo contrario, no la rechazamos.

Para el ejemplo, los resultados de la regresión son los siguientes:

$$\hat{Y}_i = 166.467 + 19.333X_i \tag{13.4.13}$$

donde $Y$ es el costo total y $X$ es la producción. Los errores estándar para esta regresión ya están en la tabla 13.1.

Cuando se hace la regresión con los residuos de (13.4.13), como se acaba de sugerir en el paso 3, obtenemos los siguientes resultados:

$$\begin{aligned}\hat{u}_i &= -24.7 \quad + 43.5443X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ \text{ee} &= \quad (6.375) \quad (4.779) \quad (0.986) \quad (0.059) \\ &\qquad R^2 = 0.9896\end{aligned} \tag{13.4.14}$$

Aunque el tamaño de la muestra es de 10, es decir, no es grande, sólo para ilustrar el mecanismo ML, obtenemos $nR^2 = (10)(0.9896) = 9.896$. De la tabla ji cuadrada observamos que, para 2 gl, el valor ji cuadrada crítico a 1% es alrededor de 9.21. Por consiguiente, el valor observado de 9.896 es significativo en el nivel de 1% y la conclusión sería rechazar la regresión restringida (es decir, la función lineal de costos). Con base en la prueba RESET de Ramsey llegamos a una conclusión similar.

## 13.5 Errores de medición

Todo el tiempo hemos supuesto implícitamente que las mediciones de la variable dependiente $Y$ y de las variables explicativas, las $X$, se realizan sin error. Así, en la regresión del gasto de consumo sobre el ingreso y la riqueza de las unidades familiares suponemos que la información sobre estas variables es "precisa"; que no se trata de estimaciones supuestas, extrapolaciones, interpolaciones o aproximaciones realizadas en forma sistemática, como la aproximación a la centésima de dólar más cercana y así sucesivamente. Por desgracia, este ideal no se cumple en la práctica por diversas razones, como errores de no respuesta, en los informes y de computación. Cualesquiera que sean las razones, el error de medición es un problema en potencia complicado, pues constituye aún otro ejemplo de sesgo de especificación con las consecuencias que veremos en seguida.

### Errores de medición en la variable dependiente *Y*

Considere el siguiente modelo:

$$Y_i^* = \alpha + \beta X_i + u_i \tag{13.5.1}$$

[26] R.F. Engle, "A General Approach to Lagrangian Multiplier Model Diagnostics", *Journal of Econometrics,* vol. 20, 1982, pp. 83-104.

donde $Y_i^* =$ gasto de consumo permanente[27]
$X_i =$ ingreso actual
$u_i =$ término de perturbación estocástico

Como $Y_i^*$ no puede medirse directamente, podemos utilizar una variable de gasto observable $Y_i$ tal que

$$Y_i = Y_i^* + \varepsilon_i \tag{13.5.2}$$

donde $\varepsilon_i$ denota los errores de medición en $Y_i^*$. Por consiguiente, en lugar de estimar (13.5.1), estimamos

$$\begin{aligned} Y_i &= (\alpha + \beta X_i + u_i) + \varepsilon_i \\ &= \alpha + \beta X_i + (u_i + \varepsilon_i) \\ &= \alpha + \beta X_i + v_i \end{aligned} \tag{13.5.3}$$

donde $v_i = u_i + \varepsilon_i$ es un término de error compuesto, que contiene el término de perturbación poblacional (el cual puede llamarse *término de error ecuacional*) y el término de error de medición.

Por simplicidad, suponga que $E(u_i) = E(\varepsilon_i) = 0$, $\text{cov}(X_i, u_i) = 0$ (el supuesto de la regresión lineal clásica) y la $\text{cov}(X_i, \varepsilon_i) = 0$; es decir, los errores de medición en $Y_i^*$ no están correlacionados con $X_i$ y la $\text{cov}(u_i, \varepsilon_i) = 0$; es decir, el error ecuacional y el error de medición no están correlacionados. Con estos supuestos, vemos que la $\beta$ estimada de (13.5.1) o (13.5.3) será un estimador insesgado de la verdadera $\beta$ (véase el ejercicio 13.7); es decir, los errores de medición en la variable dependiente $Y$ no destruyen la propiedad de insesgamiento de los estimadores de MCO. Sin embargo, las varianzas y los errores estándar de la $\beta$ estimada de (13.5.1) y (13.5.3) serán diferentes porque, con las fórmulas usuales (véase el capítulo 3), obtenemos

$$\text{Modelo (13.5.1):} \qquad \text{var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} \tag{13.5.4}$$

$$\begin{aligned} \text{Modelo (13.5.3):} \qquad \text{var}(\hat{\beta}) &= \frac{\sigma_v^2}{\sum x_i^2} \\ &= \frac{\sigma_u^2 + \sigma_\varepsilon^2}{\sum x_i^2} \end{aligned} \tag{13.5.5}$$

Obviamente, la última varianza es más grande que la primera.[28] Por tanto, **aunque los errores de medición en la variable dependiente aún producen estimaciones insesgadas de los parámetros y de sus varianzas, las varianzas estimadas ahora son más grandes que cuando no existen tales errores de medición.**

## Errores de medición en la variable explicativa *X*

Suponga ahora que, en lugar de (13.5.1), tenemos el siguiente modelo:

$$Y_i = \alpha + \beta X_i^* + u_i \tag{13.5.6}$$

donde $Y_i =$ gasto de consumo actual
$X_i^* =$ ingreso permanente
$u_i =$ término de perturbación (error ecuacional)

[27] Esta frase se atribuye a Milton Friedman. Véase también el ejercicio 13.8.

[28] Sin embargo, observe que esta varianza es aún insesgada porque, en las condiciones establecidas, el término de error compuesto $v_i = u_i + \varepsilon_i$ aún satisface los supuestos en los cuales se basa el método de mínimos cuadrados.

Suponga que en lugar de $X_i^*$, observamos

$$X_i = X_i^* + w_i \tag{13.5.7}$$

donde $w_i$ representa los errores de medición en $X_i^*$. Por consiguiente, en lugar de estimar (13.5.6), estimamos

$$\begin{aligned} Y_i &= \alpha + \beta(X_i - w_i) + u_i \\ &= \alpha + \beta X_i + (u_i - \beta w_i) \\ &= \alpha + \beta X_i + z_i \end{aligned} \tag{13.5.8}$$

donde $z_i = u_i - \beta w_i$, una composición de errores ecuacional y de medición.

Ahora bien, aunque supongamos que $w_i$ tiene media cero, es serialmente independiente y no está correlacionado con $u_i$, no podemos suponer todavía que el término de error compuesto $z_i$ es independiente de la variable explicativa $X_i$ porque [suponiendo que $E[z_i] = 0$]

$$\begin{aligned} \operatorname{cov}(z_i, X_i) &= E[z_i - E(z_i)][X_i - E(X_i)] \\ &= E(u_i - \beta w_i)(w_i) \qquad \text{con (13.5.7)} \\ &= E\left(-\beta w_i^2\right) \\ &= -\beta\sigma_w^2 \end{aligned} \tag{13.5.9}$$

Así, la variable explicativa y el término de error en (13.5.8) están correlacionados, lo cual viola el supuesto básico del modelo clásico de regresión lineal de que la variable explicativa no está correlacionada con el término de perturbación estocástico. Si se viola este supuesto, puede demostrarse que los *estimadores de MCO no solamente están sesgados, sino que son también inconsistentes, es decir, permanecen sesgados aunque el tamaño de la muestra, n, aumente indefinidamente*.[29]

Para el modelo (13.5.8), se demuestra en el apéndice 13A, sección 13A.3, que

$$\text{plím}\,\hat{\beta} = \beta\left[\frac{1}{1 + \sigma_w^2/\sigma_{X^*}^2}\right] \tag{13.5.10}$$

donde $\sigma_w^2$ y $\sigma_{X^*}^2$ son las varianzas de $w_i$ y de $X^*$, respectivamente, y donde plím$\hat{\beta}$ significa el límite en probabilidad de $\beta$.

Como se espera que el término entre corchetes sea menor que 1 (¿por qué?), (13.5.10) indica que, aunque el tamaño de la muestra aumente indefinidamente, $\hat{\beta}$ no convergirá hacia $\beta$. De hecho, si se supone que $\beta$ es positivo, $\hat{\beta}$ subestimará a $\beta$, es decir, es sesgado hacia cero. Por supuesto, si no hay errores de medición en $X$ (es decir, $\sigma_w^2 = 0$), $\hat{\beta}$ servirá como estimador consistente de $\beta$.

Por tanto, los errores de medición constituyen un grave problema cuando están presentes en la(s) variable(s) explicativa(s) porque su presencia hace imposible la estimación consistente de los parámetros. Por supuesto, como vimos, si éstos están presentes sólo en la variable dependiente, los estimadores permanecen insesgados y, por ende, son por igual consistentes. Si los errores de medición están presentes en las variables explicativas, ¿cuál es la solución? La respuesta no es fácil. En un extremo, podemos suponer que si $\sigma_w^2$ es pequeña comparada con $\sigma_{X^*}^2$, para todos los fines prácticos podemos suponer "que no existe" el problema y proceder con la estimación usual

[29] Como se demuestra en el **apéndice A**, $\hat{\beta}$ es un estimador consistente de $\beta$ pues, a medida que $n$ aumenta indefinidamente, la distribución muestral de $\hat{\beta}$ tenderá al verdadero $\beta$. En términos técnicos, esto se plantea así: $\text{plím}_{n\to\infty}\hat{\beta} = \beta$. Como se anota en el **apéndice A**, la consistencia es una propiedad de muestras grandes y a menudo se utiliza para estudiar el comportamiento de un estimador cuando no pueden determinarse sus propiedades finitas o de muestras pequeñas (por ejemplo, insesgamiento).

por MCO. Desde luego, el tropiezo aquí es que no es posible observar o medir $\sigma_w^2$ ni $\sigma_{X^*}^2$ fácilmente, y por consiguiente, no hay forma de juzgar sus magnitudes relativas.

Otro remedio sugerido son las **variables instrumentales** o **representantes** (**proxy**) que, aunque están muy correlacionadas con las variables $X$ originales, no están correlacionadas con los términos de error ecuacional y de medición (es decir, $u_i$ y $w_i$). Si es posible encontrar tales variables representantes, también lo es obtener una estimación consistente de $\beta$. Pero es mucho más fácil hablar sobre esta labor que hacerla. En la práctica, no es fácil encontrar buenas variables representantes; con frecuencia estamos en una situación de inconformidad sobre el mal clima sin ser capaces de hacer mucho al respecto. Además, no es fácil saber si la variable instrumental seleccionada es en realidad independiente de los términos de error $u_i$ y $w_i$.

En la teoría hay otras sugerencias para resolver el problema.[30] Pero la mayoría es específica de cada situación y sus supuestos son restrictivos. En realidad no hay respuesta satisfactoria al problema de los errores de medición. Por esto es tan crucial que la medición de los datos sea lo más precisa posible.

**EJEMPLO 13.2**
*Un ejemplo*

Concluimos esta sección con un ejemplo construido para resaltar los puntos anteriores.

La tabla 13.2 proporciona información hipotética sobre el gasto de consumo verdadero $Y^*$, el ingreso verdadero $X^*$, el consumo medido $Y$ y el ingreso medido $X$. La tabla también explica la forma como se midieron estas variables.[31]

**Errores de medición sólo en la variable dependiente *Y*.** Con base en esta información, la verdadera función de consumo es

$$\begin{aligned} \hat{Y}_i^* &= 25.00 \quad + \quad 0.6000X_i^* \\ &\quad (10.477) \qquad (0.0584) \\ t &= \ (2.3861) \quad (10.276) \\ &\qquad\qquad R^2 = 0.9296 \end{aligned} \tag{13.5.11}$$

**TABLA 13.2**
**Información hipotética sobre *Y** (verdadero gasto de consumo), *X** (verdadero ingreso), *Y* (gasto de consumo medido) y *X* (ingreso medido). Todas las cifras están en dólares**

| $Y^*$ | $X^*$ | $Y$ | $X$ | $\varepsilon$ | $w$ | $u$ |
|---|---|---|---|---|---|---|
| 75.4666 | 80.00 | 67.6011 | 80.0940 | −7.8655 | 0.0940 | 2.4666 |
| 74.9801 | 100.00 | 75.4438 | 91.5721 | 0.4636 | −8.4279 | −10.0199 |
| 102.8242 | 120.00 | 109.6956 | 112.1406 | 6.8714 | 2.1406 | 5.8242 |
| 125.7651 | 140.00 | 129.4159 | 145.5969 | 3.6509 | 5.5969 | 16.7651 |
| 106.5035 | 160.00 | 104.2388 | 168.5579 | −2.2647 | 8.5579 | −14.4965 |
| 131.4318 | 180.00 | 125.8319 | 171.4793 | −5.5999 | −8.5207 | −1.5682 |
| 149.3693 | 200.00 | 153.9926 | 203.5366 | 4.6233 | 3.5366 | 4.3693 |
| 143.8628 | 220.00 | 152.9208 | 222.8533 | 9.0579 | 2.8533 | −13.1372 |
| 177.5218 | 240.00 | 176.3344 | 232.9879 | −1.1874 | −7.0120 | 8.5218 |
| 182.2748 | 260.00 | 174.5252 | 261.1813 | −7.7496 | 1.1813 | 1.2748 |

*Nota:* Se supone que los datos sobre $X^*$ están dados. En la derivación de las demás variables, los supuestos fueron los siguientes: 1) $E(u_i) = E(\varepsilon_i) = E(w_i) = 0$; 2) cov $(X, u) =$ cov $(X, \varepsilon) =$ cov $(u, \varepsilon) =$ cov $(w, u) =$ cov $(\varepsilon, w) = 0$; 3) $\sigma_u^2 = 100$, $\sigma_s^2 = 36$, y $\sigma_w^2 = 36$; y 4) $Y_i^* = 25 + 0.6X_i^* + u_i$, $Y_i = Y_i^* + \varepsilon_i$ y $X_i = X_i^* + w_i$.

(*continúa*)

[30] Véase Thomas B. Fomby, R. Carter Hill y Stanley R. Johnson, *Advanced Econometric Methods,* Springer-Verlag, Nueva York, 1984, pp. 273-277. Véase también Kennedy, *op. cit.*, pp. 138-140, para un análisis sobre regresión ponderada y variables instrumentales. También G.S. Maddala, *Introduction to Econometrics,* 3a. ed., John Wiley & Sons, Nueva York, 2001, pp. 437-462, y Quirino Paris, "Robust Estimators of Errors-in-Variables Models: Part I", documento de trabajo núm. 04-007, 200, Departamento de Economía Agrícola y de Recursos, Universidad de California en Davis, agosto de 2004.

[31] El autor agradece a Kenneth J. White la elaboración de este ejemplo. Véase su *Computer Handbook Using SHAZAM*, para utilizarse con Damodar Gujarati, *Basic Econometrics*, septiembre de 1985, pp. 117-121.

**EJEMPLO 13.2**
(*continuación*)

mientras que si utilizamos $Y_i$ en lugar de $Y_i^*$, obtenemos

$$\begin{aligned} \hat{Y}_i &= 25.00 \quad + \; 0.6000X_i^* \\ &\quad (12.218) \quad (0.0681) \\ t &= \; (2.0461) \quad (8.8118) \\ &\qquad\qquad R^2 = 0.9066 \end{aligned} \tag{13.5.12}$$

Como indican estos resultados y de acuerdo con la teoría, los coeficientes estimados continúan siendo iguales. El único efecto de los errores de medición en la variable dependiente es que los errores estándar estimados de los coeficientes tienden a ser más grandes [véase (13.5.5)], lo cual se aprecia con claridad en (13.5.12). A propósito, observe que los coeficientes de regresión en (13.5.11) y (13.5.12) son los mismos porque la muestra se generó para cumplir con los supuestos del modelo de errores de medición.

**Errores de medición en *X*.** Sabemos que la regresión verdadera es (13.5.11). Suponga ahora que en lugar de $X_i^*$ utilizamos $X_i$. (*Nota:* En realidad, $X_i^*$ pocas veces es observable.) Los resultados de la regresión son los siguientes:

$$\begin{aligned} \hat{Y}_i^* &= 25.992 \quad + \; 0.5942X_i \\ &\quad (11.0810) \quad (0.0617) \\ t &= \; (2.3457) \quad (9.6270) \\ &\qquad\qquad R^2 = 0.9205 \end{aligned} \tag{13.5.13}$$

Estos resultados están de acuerdo con la teoría: cuando hay errores de medición en la(s) variable(s) explicativa(s), los coeficientes estimados están sesgados. Por fortuna, en este ejemplo el sesgo es relativamente pequeño; de (13.5.10) es evidente que el sesgo depende de $\sigma_w^2/\sigma_{X^*}^2$, y en la generación de la información supusimos que $\sigma_w^2 = 36$ y $\sigma_{X^*}^2 = 3\,667$, con lo que redujimos el factor de sesgo, alrededor de 0.98% (= 36/3 667).

Dejamos al lector averiguar lo que sucede cuando hay errores de medición en *Y* y en *X*, es decir, cuando efectuamos la regresión de $Y_i$ sobre $X_i$ en lugar de hacerla de $Y_i^*$ sobre $X_i^*$ (véase el ejercicio 13.23).

## 13.6 Especificación incorrecta del término de error estocástico

Un problema común de los investigadores es la especificación del término de error $u_i$, que ingresa en el modelo de regresión. Como el término de error no se puede observar de manera directa, no hay una forma sencilla de determinar la forma en que ingresa en el modelo. A fin de ver lo anterior, considere los modelos de (13.2.8) y (13.2.9). Por simplicidad de la exposición, supusimos que no hay intercepto en el modelo. Además, supondremos que $u_i$ en (13.2.8) es tal que $\ln u_i$ satisface los supuestos característicos de MCO.

Si suponemos que (13.2.8) es el modelo "correcto" pero estimamos (13.2.9), ¿cuáles son las consecuencias? En el apéndice 13.A, sección 13A.4, se muestra que si $\ln u_i \sim N(0, \sigma^2)$, entonces

$$u_i \sim \log \text{normal} \left[e^{\sigma^2/2},\ e^{\sigma^2}\left(e^{\sigma^2} - 1\right)\right] \tag{13.6.1}$$

como resultado:

$$E(\hat{\alpha}) = \beta e^{\sigma^2/2} \tag{13.6.2}$$

donde *e* es la base del logaritmo natural.

Como se aprecia, $\hat{\alpha}$ es un estimador sesgado, pues su valor promedio no es igual a la verdadera $\beta$.

Veremos más respecto de la especificación del término de error estocástico en el capítulo sobre los modelos de regresión no lineales en los parámetros.

## 13.7 Modelos anidados y no anidados

Al efectuar la prueba de especificación, es útil diferenciar entre **modelos anidados** y **no anidados**. Para distinguirlos, considere los siguientes modelos:

$$\text{Modelo A:} \qquad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

$$\text{Modelo B:} \qquad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Decimos que el modelo B está anidado en el modelo A porque es un caso especial del modelo A: si estimamos el modelo A y probamos la hipótesis de que $\beta_4 = \beta_5 = 0$ y no se rechaza con base en la prueba $F$,[32] el modelo A se reduce al modelo B. Si añadimos la variable $X_4$ al modelo B, el A se reducirá al B, si $\beta_5$ es cero; en este caso aplicaremos la prueba $t$ a la hipótesis de que el coeficiente de $X_5$ es cero.

Sin llamarlas de ese modo, las pruebas de error de especificación que acabamos de analizar y la prueba $F$ restringida que vimos en el capítulo 8 son en esencia pruebas de hipótesis anidadas.

Ahora considere los siguientes modelos:

$$\text{Modelo C:} \qquad Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$$

$$\text{Modelo D:} \qquad Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$$

donde las $X$ y las $Z$ son variables distintas. Decimos que los modelos C y D son **no anidados** porque no puede derivarse uno como caso especial del otro. En economía, como en otras ciencias, más de una teoría puede explicar un fenómeno. Por tanto, los monetaristas pueden destacar la función del dinero al explicar los cambios del PIB, en tanto que los keynesianos pueden explicarlos mediante las variaciones en el gasto gubernamental.

Debe notarse que se puede permitir que los modelos C y D contengan regresoras comunes a ambos. Por ejemplo, $X_3$ puede incluirse en D, y $Z_2$ en C. Aun así, estos modelos son no anidados, pues el modelo C no contiene a $Z_3$, y el modelo D no contiene a $X_2$.

Aunque se encuentren las mismas variables en el modelo, por la forma funcional pueden ser dos modelos no anidados. Por ejemplo, considere el modelo:

$$\text{Modelo E:} \qquad Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$$

Los modelos D y E son no anidados, pues no se puede derivar uno como caso especial del otro.

Como ya vimos las pruebas de modelos anidados (pruebas $t$ y $F$), en la siguiente sección analizaremos algunas pruebas para los modelos no anidados, antes llamados errores de especificación incorrecta del modelo.

[32] De manera más general, es posible utilizar la prueba de la razón de verosimilitud o la de Wald, o bien la prueba del multiplicador de Lagrange, que analizamos brevemente en el capítulo 8.

## 13.8 Pruebas de hipótesis no anidadas

De acuerdo con Harvey,[33] existen dos métodos para probar hipótesis no anidadas: 1) el **método de discriminación**, en donde dados dos o más modelos rivales, uno elige un modelo con base en criterios de bondad de ajuste, y 2) el **método de discernimiento** (en la terminología de este texto), en donde al investigar un modelo, se toma en cuenta la información proporcionada por otros modelos. Estudiaremos brevemente ambos métodos.

### Método de discriminación

Considere los modelos C y D anteriores. Como ambos tienen la misma variable dependiente, podemos elegir entre dos (o más) modelos con base en algún criterio de bondad de ajuste, como $R^2$ o $R^2$ ajustada, ya analizado. Pero tenga en cuenta que al comparar dos o más modelos, la regresada debe ser la misma. Además de estos criterios, hay otros también comunes. Entre ellos están el **criterio de información de Akaike (CIA), el criterio de información de Schwarz (CIS)** y el **criterio $C_p$ de Mallows**. Los estudiaremos en la sección 13.9. El software más moderno de estadística contiene uno o más de tales criterios intercalados en sus rutinas de regresión. En la última sección de este capítulo ilustraremos los criterios anteriores con un ejemplo ampliado. Con base en uno o más de tales criterios seleccionamos finalmente un modelo con la máxima $\bar{R}^2$, o el valor más bajo del CIA o del CIS, etcétera.

### Método de discernimiento

*La prueba F no anidada o la prueba F incluyente*

Considere los modelos C y D presentados en la sección 3.7. ¿Cómo elegir entre ambos modelos? Para este propósito, suponga que estimamos el siguiente modelo anidado o *híbrido*:

$$\text{Modelo F:} \qquad Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 Z_{2i} + \lambda_5 Z_{3i} + u_i$$

Observe que el modelo F *anida* o *incluy*e a los modelos C y D. Pero note que C no está anidado en D, y que éste no está anidado en C, por lo que no son modelos anidados.

Ahora bien, si el modelo C es correcto, $\lambda_4 = \lambda_5 = 0$, en tanto que D es correcto si $\lambda_2 = \lambda_3 = 0$. Esta prueba se efectúa mediante la prueba $F$ usual, de aquí que se le conozca como prueba $F$ no anidada.

Sin embargo, surgen problemas con este procedimiento de prueba. En *primer* lugar, si las $X$ y las $Z$ están demasiado correlacionadas, entonces —como vimos en el capítulo de multicolinealidad— es muy probable que una o más de las $\lambda$ sean en lo individual estadísticamente insignificantes, aunque con base en la prueba $F$ podamos rechazar la hipótesis de que todos los coeficientes de pendientes sean simultáneamente nulos. En este caso, no hay forma de decidir si el modelo C o el D es el correcto. En *segundo* lugar, existe otro problema. Suponga que elegimos el modelo C como *hipótesis de referencia* o modelo, y descubrimos que todos sus coeficientes son significativos. Ahora agregamos $Z_2$ o $Z_3$, o ambas, al modelo y tenemos que, al utilizar la prueba $F$, su contribución incremental a la suma de cuadrados explicada (SCE) es estadísticamente insignificante. Por tanto, elegimos el modelo C.

Pero suponga que elegimos el modelo D como referencia y encontramos que todos sus coeficientes son estadísticamente significativos. Pero cuando agregamos $X_2$ o $X_3$, o ambas, a este modelo, de nuevo observamos que, con la prueba $F$, su contribución incremental a la SCE es insignificante. En consecuencia, habríamos elegido el modelo D como el correcto. Por tanto, "la elección de la hipótesis de referencia puede determinar el resultado de la elección del modelo",[34] sobre todo si hay una gran multicolinealidad en las regresoras rivales. Por *último*, el modelo $F$, artificialmente anidado, quizá no tenga ningún significado económico.

[33] Andrew Harvey, *The Econometric Analysis of Time Series,* 2a. ed., The MIT Press, Cambridge, Massachusetts, 1990, cap. 5.

[34] Thomas B. Fomby, R. Carter Hill y Stanley R. Johnson, *Advanced Econometric Methods,* Springer Verlag, Nueva York, 1984, p. 416.

**EJEMPLO 13.3**
*Un ejemplo ilustrativo: El modelo St. Louis*

Para determinar si los cambios en el PIB nominal se explican por las variaciones en la oferta de dinero (monetarismo) o por los cambios en el gasto gubernamental (keynesianismo), consideramos los siguientes modelos:

$$\dot{Y}_t = \alpha + \beta_0 \dot{M}_t + \beta_1 \dot{M}_{t-1} + \beta_2 \dot{M}_{t-2} + \beta_3 \dot{M}_{t-3} + \beta_4 \dot{M}_{t-4} + u_{1t}$$

$$= \alpha + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + u_{1t} \tag{13.8.1}$$

$$\dot{Y}_t = \gamma + \lambda_0 \dot{E}_t + \lambda_1 \dot{E}_{t-1} + \lambda_2 \dot{E}_{t-2} + \lambda_3 \dot{E}_{t-3} + \lambda_4 \dot{E}_{t-4} + u_{2t}$$

$$= \gamma + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{2t} \tag{13.8.2}$$

donde $\dot{Y}_t$ = tasa de crecimiento en el PIB nominal para el tiempo $t$
$\dot{M}_t$ = tasa de crecimiento en la oferta de dinero (versión $M_1$) en el tiempo $t$
$\dot{E}_t$ = tasa de crecimiento con una plena o alta utilización del gasto gubernamental en el tiempo $t$

Por cierto, observe que (13.8.1) y (13.8.2) son ejemplos de modelos de rezago distribuido, tema que analizaremos a profundidad en el capítulo 17. Por el momento, simplemente note que el efecto de una unidad de cambio en la oferta de dinero o en el gasto gubernamental sobre el PIB se distribuye a lo largo del tiempo y no es instantáneo.

Como sería difícil, *a priori*, decidir entre los dos modelos rivales, mezclaremos ambos como se muestra a continuación:

$$\dot{Y}_t = \text{constante} + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{3t} \tag{13.8.3}$$

Este modelo anidado es una forma en la que se ha expresado y estimado el famoso modelo (del Banco de la Reserva Federal) de St. Louis, un banco de la escuela monetarista. Sus resultados para el periodo del primer trimestre de 1953 al cuarto de 1976 para Estados Unidos son los siguientes (las razones $t$ están entre paréntesis):[35]

| Coeficiente | Estimado | | Coeficiente | Estimado | |
|---|---|---|---|---|---|
| $\beta_0$ | 0.40 | (2.96) | $\lambda_0$ | 0.08 | (2.26) |
| $\beta_1$ | 0.41 | (5.26) | $\lambda_1$ | 0.06 | (2.52) |
| $\beta_2$ | 0.25 | (2.14) | $\lambda_2$ | 0.00 | (0.02) |
| $\beta_3$ | 0.06 | (0.71) | $\lambda_3$ | −0.06 | (−2.20) |
| $\beta_4$ | −0.05 | (−0.37) | $\lambda_4$ | −0.07 | (−1.83) |
| $\sum_{i=0}^{4} \beta_i$ | 1.06 | (5.59) | $\sum_{i=0}^{4} \lambda_i$ | 0.03 | (0.40) |

$R^2 = 0.40$
$d = 1.78$ **(13.8.4)**

¿Qué indican estos resultados en lo que concierne a la superioridad de un modelo respecto del otro? Si nos atenemos al efecto acumulativo de una unidad de cambio en $\dot{M}$ y $\dot{E}$ sobre $\dot{Y}$, obtenemos respectivamente $\sum_{i=0}^{4} \beta_i = 1.06$ y $\sum_{i=0}^{4} \lambda_i = 0.03$, con el primero estadísticamente significativo y el último no. Esta comparación apoyaría la afirmación monetarista de que los cambios en la oferta de dinero son los que determinan las variaciones en el PIB (nominal). Se deja como ejercicio para el lector evaluar en forma crítica esta afirmación.

[35] Véase Keith M. Carlson, "Does the St. Louis Equation Now Believe in Fiscal Policy?", *Review, Federal Reserve Bank of St. Louis,* vol. 60, núm. 2, febrero de 1978, p. 17, tabla IV.

### *La prueba J de Davidson-MacKinnon*[36]

En vista de los problemas que acabamos de mencionar en el procedimiento de prueba *F* no anidado, se han sugerido otras opciones. Una es la **prueba *J* de Davidson-MacKinnon**. Para ilustrarla, suponga que deseamos comparar la hipótesis o modelo C con la hipótesis o modelo D. La prueba *J* procede de la siguiente forma:

1. Estimamos el modelo D y de él obtenemos los valores *Y* estimados, $\hat{Y}_i^D$.

2. Agregamos el valor *Y* pronosticado en el paso 1 como una regresora adicional al modelo C y estimamos el siguiente modelo:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 \hat{Y}_i^D + u_i \qquad \textbf{(13.8.5)}$$

donde los valores $\hat{Y}_i^D$ se obtienen del paso 1. Este modelo es un ejemplo del **principio de inclusión**, como en la metodología de Hendry.

3. Con la prueba *t*, se prueba la hipótesis de que $\alpha_4 = 0$.

4. Si no se rechaza la hipótesis de que $\alpha_4 = 0$, podemos aceptar (es decir, no se rechaza) el modelo C como el verdadero modelo, pues $\hat{Y}_i^D$, incluida en (13.8.5), que representa la influencia de las variables no consideradas en el modelo C, no tiene un poder explicativo adicional más allá de lo que contribuye el modelo C. En otras palabras, el modelo C *incluye* al modelo D, en el sentido de que este último no contiene ninguna información adicional que mejore el desempeño de C. Por el mismo tenor, si se rechaza la hipótesis nula, el modelo C no puede ser el verdadero (¿por qué?).

5. Ahora cambiamos los papeles de las hipótesis, o de los modelos C y D. Estimamos primero el modelo C, con los valores *Y* estimados de este modelo como regresoras en (13.8.5), repetimos el paso 4 y decidimos si preferimos o no el modelo D respecto del C. De manera más específica, estimamos el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 \hat{Y}_i^C + u_i \qquad \textbf{(13.8.6)}$$

donde $\hat{Y}_i^C$ son los valores *Y* estimados del modelo C. Ahora probamos la hipótesis de que $\beta_4 = 0$. Si no se rechaza esta hipótesis, elegimos el modelo D en vez del C. Si se rechaza la hipótesis de que $\beta_4 = 0$, entonces preferiremos C en vez de D, pues este último no tiene un mejor desempeño que C.

Aunque resulta intuitivamente llamativa, la prueba *J* presenta algunos problemas. Como las pruebas dadas en (13.8.5) y (13.8.6) se realizan de manera independiente, tenemos los siguientes resultados probables:

| | **Hipótesis: $\alpha_4 = 0$** | |
|---|---|---|
| **Hipótesis: $\beta_4 = 0$** | **No se rechaza** | **Se rechaza** |
| **No se rechaza** | Se aceptan C y D | Se acepta D, se rechaza C |
| **Se rechaza** | Se acepta C, se rechaza D | Se rechazan C y D |

Como muestra la tabla, no podremos dar una respuesta contundente si el procedimiento de prueba *J* conduce a la aceptación o rechazo de ambos modelos. En caso de que ambos se rechacen, ningún modelo explica el comportamiento de *Y*. De igual forma, si ambos se aceptan, como observa Kmenta, "los datos al parecer no son lo bastante ricos para discriminar entre las dos hipótesis [modelos]".[37]

---

[36] R. Davidson y J.G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica,* vol. 49, 1981, pp. 781–793.

[37] Jan Kmenta, *op. cit.,* p. 597.

Otro problema con la prueba $J$ es que cuando se utiliza el estadístico $t$ para probar la significancia de la variable $Y$ estimada en los modelos (13.8.5) y (13.8.6), el estadístico $t$ tiene la distribución normal estándar sólo de manera asintótica, es decir, para muestras grandes. Por consiguiente, la prueba $J$ quizá no sea muy poderosa (en el sentido estadístico) para muestras pequeñas, pues tiende a rechazar la hipótesis o el modelo verdadero con una frecuencia mayor de la que debería.

**EJEMPLO 13.4**
*Gasto de consumo personal e ingreso personal disponible*

Para ilustrar la prueba $J$, considere los datos de la tabla 13.3, la cual proporciona el gasto de consumo personal *per cápita* (GCPP) y el ingreso personal disponible *per cápita* (IPDP), ambos en dólares de 2008, en Estados Unidos de 1970 a 2005. Ahora considere los siguientes modelos rivales:

$$\text{Modelo A:} \quad \text{GCPP}_t = \alpha_1 + \alpha_2 \text{IPDP}_t + \alpha_3 \text{IPDP}_{t-1} + u_t \qquad \textbf{(13.8.7)}$$

$$\text{Modelo B:} \quad \text{GCPP}_t = \beta_1 + \beta_2 \text{IPDP}_t + \beta_3 \text{GCPP}_{t-1} + u_t \qquad \textbf{(13.8.8)}$$

El modelo A establece que el GCPP depende del IPDP en el periodo actual y previo; este modelo es un ejemplo de **modelo de rezago distribuido** (véase el capítulo 17). El modelo B postula que el GCPP depende del IPDP actual y del GCPP del periodo anterior; este modelo representa el **modelo autorregresivo** (véase el capítulo 17). La razón para introducir el valor rezagado del GCPP en este modelo es reflejar la inercia o persistencia del hábito.

Los resultados de estimar estos modelos por separado fueron los siguientes:

$$\text{Modelo A:} \quad \widehat{\text{GCPP}}_t = -606.6347 + 0.6170\ \text{IPDP}_t + 0.3530\ \text{IPDP}_{t-1}$$
$$t = (-3.8334) \quad (2.5706) \quad (1.4377) \qquad \textbf{(13.8.9)}$$
$$R^2 = 0.9983 \qquad d = 0.2161$$

$$\text{Modelo B:} \quad \widehat{\text{GCPP}}_t = 76.8947 + 0.2074\ \text{IPDP}_t + 0.8104\ \text{GCPP}_{t-1}$$
$$t = (0.7256) \quad (2.6734) \quad (9.7343) \qquad \textbf{(13.8.10)}$$
$$R^2 = 0.9996 \qquad d = 0.9732$$

**TABLA 13.3**
**Gasto de consumo personal *per cápita* (GCPP) e ingreso personal disponible *per cápita* (IPDP) en Estados Unidos, 1970-2005**

Fuente: *Economic Report of the President*, 2007.

| Año | GCPP | IPDP | Año | GCPP | IPDP |
|---|---|---|---|---|---|
| 1970 | 3 162 | 3 587 | 1988 | 13 685 | 15 297 |
| 1971 | 3 379 | 3 860 | 1989 | 14 546 | 16 257 |
| 1972 | 3 671 | 4 140 | 1990 | 15 349 | 17 131 |
| 1973 | 4 022 | 4 616 | 1991 | 15 722 | 17 609 |
| 1974 | 4 364 | 5 010 | 1992 | 16 485 | 18 494 |
| 1975 | 4 789 | 5 498 | 1993 | 17 204 | 18 872 |
| 1976 | 5 282 | 5 972 | 1994 | 18 004 | 19 555 |
| 1977 | 5 804 | 6 517 | 1995 | 18 665 | 20 287 |
| 1978 | 6 417 | 7 224 | 1996 | 19 490 | 21 091 |
| 1979 | 7 073 | 7 967 | 1997 | 20 323 | 21 940 |
| 1980 | 7 716 | 8 822 | 1998 | 21 291 | 23 161 |
| 1981 | 8 439 | 9 765 | 1999 | 22 491 | 23 968 |
| 1982 | 8 945 | 10 426 | 2000 | 23 862 | 25 472 |
| 1983 | 9 775 | 11 131 | 2001 | 24 722 | 26 235 |
| 1984 | 10 589 | 12 319 | 2002 | 25 501 | 27 164 |
| 1985 | 11 406 | 13 037 | 2003 | 26 463 | 28 039 |
| 1986 | 12 048 | 13 649 | 2004 | 27 937 | 29 536 |
| 1987 | 12 766 | 14 241 | 2005 | 29 468 | 30 458 |

(*continúa*)

**EJEMPLO 13.4** (*continuación*)

Si se tuviese que elegir entre estos dos modelos con base en el método de discriminación, según el criterio $R^2$, quizá se elegiría el modelo B (13.8.10) porque es un poco más alto que el A (13.8.9). Además, en el modelo B (13.8.10) ambas variables son estadísticamente significativas en lo individual, en tanto que en el A (13.8.9) sólo el IPDP actual es estadísticamente significativo (aunque puede haber un problema de colinealidad). Sin embargo, para efectos predictivos no existe mucha diferencia entre los dos valores estimados de $R^2$.

Para aplicar la prueba *J*, suponga que el modelo A es la hipótesis nula, es decir, el modelo mantenido, y el modelo B es la hipótesis alternativa. Siguiendo los pasos de la prueba *J* analizados antes, se utilizan los valores estimados del GCPP del modelo (13.8.10) como una regresora incondicional en el modelo A, con el siguiente resultado:

$$\widehat{\text{GCPP}}_t = -35.17 + 0.2762\,\text{IPDP}_t - 0.5141\,\text{IPDP}_{t-1} + 1.2351\,\widehat{\text{GCPP}}_t^B$$
$$t = (-0.43) \quad (2.60) \quad (-4.05) \quad (12.06) \tag{13.8.11}$$
$$R^2 = 1.00 \quad d = 1.5205$$

donde $\widehat{\text{GCPP}}_t^B$ en el miembro derecho de (13.8.11) representa los valores estimados GCPP del modelo B (13.8.10). Como el coeficiente de esta variable es estadísticamente significativo con un estadístico *t* muy alto de 12.06, según el procedimiento de la prueba *J* se tiene que rechazar el modelo A y aceptar el B.

Ahora supondremos que el modelo B es la hipótesis mantenida y que el A es la alternativa, exactamente con el mismo procedimiento que antes, y obtenemos los siguientes resultados:

$$\widehat{\text{GCPP}}_t = -823.7 + 1.4309\,\text{IPDP}_t + 1.0009\,\text{GCPP}_{t-1} - 1.4563\,\widehat{\text{GCPP}}_t^A$$
$$t = (-3.45) \quad (4.64) \quad (12.06) \quad (-4.05) \tag{13.8.12}$$
$$R^2 = 1.00 \quad d = 1.5205$$

donde $\widehat{\text{GCPP}}_t^A$ en el miembro derecho de la ecuación (13.8.12) representa los valores estimados de GCPP del modelo A original (13.8.9). En esta regresión el coeficiente de $\widehat{\text{GCPP}}_t^A$ también es estadísticamente significativo, con un estadístico *t* de −4.05. Este resultado indica que ahora debemos preferir el modelo B en vez del A.

Todo lo anterior muestra que ningún modelo es particularmente útil para explicar el comportamiento del gasto de consumo personal *per cápita* en Estados Unidos de 1970 a 2005. Por supuesto, sólo consideramos dos modelos rivales. En realidad, bien puede haber más de dos modelos. El procedimiento de la prueba *J* puede ampliarse a la comparación de múltiples modelos, aunque así el análisis se complica con facilidad.

Este ejemplo muestra de forma muy vívida por qué el MCRL supone que el modelo de regresión del análisis está especificado de modo correcto. Obvio, resulta crucial, al desarrollar un modelo, poner especial atención al fenómeno del cual se está haciendo el modelo.

### *Otras pruebas para la selección del modelo*

La prueba *J* recién estudiada sólo es una de un grupo para seleccionar modelos: existe la **prueba Cox**, la **prueba JA**, la **prueba P**, la **prueba de inclusión Mizon-Richard** y variantes de ellas. Resulta obvio que no se espera que en este texto estudiemos estas pruebas tan especializadas, pero el lector puede consultar las referencias de las notas.[38]

[38] Véase también Badi H. Baltagi, *Econometrics,* Springer, Nueva York, 1998, pp. 209-222.

# 13.9 Criterios para la selección de modelos

En esta sección estudiaremos diversos criterios para elegir entre modelos rivales y/o comparar con propósitos de pronóstico. Aquí distinguimos entre pronóstico **dentro de la muestra** y pronóstico **fuera de la muestra**. El primero señala sobre todo cómo elegir el modelo que se ajusta a los datos de determinada muestra. El pronóstico fuera de la muestra se refiere a la forma de determinar cómo un modelo ajustado pronostica valores futuros de la regresada, dados los valores de las regresoras.

Hay diversos criterios para este fin. En particular, examinaremos los siguientes criterios: 1) $R^2$, 2) $R^2$ ajustada ($= \bar{R}^2$), 3) criterio de información Akaike (CIA), 4) criterio de información Schwarz (CIS), 5) criterio $C_p$ de Mallows y 6) pronóstico $\chi^2$ (ji cuadrada). Todos estos criterios pretenden reducir la suma de cuadrados residual (SCR) (o incrementar el valor $R^2$). Sin embargo, salvo por el primer criterio, los demás imponen un castigo por incluir un número creciente de regresoras. Por tanto, existe un dilema entre la bondad del ajuste del modelo y su complejidad (juzgada de acuerdo con el número de regresoras).

## El criterio $R^2$

Sabemos que una medida de la bondad del ajuste de un modelo de regresión es $R^2$, la cual se define como:

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} \tag{13.9.1}$$

Así definida, $R^2$ necesariamente está entre 0 y 1. Mientras más cerca esté de 1, mejor será el ajuste. Pero surgen varios problemas con $R^2$. En *primer* lugar, mide la bondad de ajuste *dentro de la muestra*, en el sentido de conocer la cercanía entre un valor $Y$ estimado y su valor real en la muestra dada. No hay garantía de que pronosticará bien las observaciones *fuera de la muestra*. En *segundo* lugar, al comparar dos o más valores de $R^2$, la variable dependiente, o regresada, debe ser la misma. En *tercer* lugar, y lo más importante, es que una $R^2$ no puede disminuir cuando se agregan más variables al modelo. Por consiguiente, existe la tentación de apostar por "maximizar $R^2$" simplemente añadiendo más variables. Por supuesto que al agregar variables se incrementa $R^2$, pero también aumenta la varianza del error de predicción.

## $R^2$ ajustada

Debido a la inconveniencia de aumentar regresoras para incrementar el valor de $R^2$, Henry Theil desarrolló la $R^2$ ajustada, denotada por $\bar{R}^2$, la cual estudiamos en el capítulo 7. Recuerde que

$$\bar{R}^2 = 1 - \frac{\text{SCE}/(n-k)}{\text{SCT}/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{13.9.2}$$

Como se ve en esta fórmula, $\bar{R}^2 \leq R^2$, lo cual muestra cómo la $R^2$ ajustada penaliza cuando se agregan más regresoras. Como observamos en el capítulo 8, a diferencia de $R^2$, la $R^2$ ajustada se incrementa sólo si el valor absoluto de $t$ de la variable añadida es mayor que 1. Así, para comparar, $\bar{R}^2$ es una mejor medida que $R^2$. Pero una vez más, tenga en cuenta que la regresada debe ser la misma para que la comparación sea válida.

## Criterio de información Akaike (CIA)

La idea de imponer una penalización por añadir regresoras al modelo se desarrolló más en el criterio CIA, el cual se define como:

$$\text{CIA} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{\text{SCR}}{\text{n}} \tag{13.9.3}$$

donde $k$ es el número de regresoras (inclusive el intercepto) y $n$ es el número de observaciones. Por conveniencia matemática (13.9.3) se expresa como

$$\ln \text{CIA} = \left(\frac{2k}{n}\right) + \ln\left(\frac{\text{SCR}}{n}\right) \tag{13.9.4}$$

donde ln CIA = el logaritmo natural de CIA y $2k/n$ = factor de penalización. Algunos libros de texto y paquetes de software definen al CIA sólo en términos de su transformada logarítmica, por lo que no es necesario escribir ln antes de CIA. Como se ve en la fórmula, CIA impone una mayor penalización que $\bar{R}^2$ por añadir regresoras. Al comparar dos o más modelos, se preferirá el que tenga el menor valor CIA. Una ventaja del CIA es que resulta útil no sólo para el desempeño de la predicción dentro de la muestra, sino también para el de la predicción fuera de la muestra de un modelo de regresión. Asimismo, es útil para los modelos anidados y no anidados. También sirve para determinar la longitud del rezago en el modelo AR($p$).

## Criterio de información Schwarz (CIS)

Con un espíritu similar al CIA, el criterio CIS se define como

$$\text{CIS} = n^{k/n}\frac{\sum \hat{u}^2}{n} = n^{k/n}\frac{\text{SCR}}{n} \tag{13.9.5}$$

o, en forma logarítmica:

$$\ln \text{CIS} = \frac{k}{n}\ln n + \ln\left(\frac{\text{SCR}}{n}\right) \tag{13.9.6}$$

donde $[(k/n) \ln n]$ es el factor de penalización. CIS impone una penalización mayor que CIA, como resulta obvio al comparar (13.9.6) con (13.9.4). Al igual que en CIA, mientras más pequeño sea el valor de CIS, mejor será el modelo. De nuevo, al igual que en CIA, CIS sirve para comparar el desempeño del pronóstico dentro de la muestra y fuera de la muestra de un modelo.

## Criterio $C_p$ de Mallows

Suponga que tenemos un modelo con $k$ regresoras, inclusive el intercepto. Sea $\hat{\sigma}^2$ el estimador de la verdadera $\sigma^2$, como siempre. Pero suponga también que sólo elegimos $p$ regresoras ($p \le k$) y obtuvimos la SCR de la regresión con esas $p$ regresoras. Sea $\text{SCR}_p$ la suma de cuadrado residual

obtenida con las $p$ regresoras. Ahora bien, C.P. Mallows elaboró el siguiente criterio para seleccionar modelos, conocido como criterio $C_p$:

$$C_p = \frac{\text{SCR}_p}{\hat{\sigma}^2} - (n - 2p) \tag{13.9.7}$$

donde $n$ es el número de observaciones.

Sabemos que $E(\hat{\sigma}^2)$ es un estimador insesgado de la verdadera $\sigma^2$. Ahora bien, si el modelo con $p$ regresoras es adecuado en lo que se refiere a que no muestra una carencia de ajuste, se puede demostrar[39] que $E(\text{SCR}_p) = (n - p)\sigma^2$. En consecuencia, es verdad aproximadamente que

$$E(C_p) \approx \frac{(n-p)\sigma^2}{\sigma^2} - (n - 2p) \approx p \tag{13.9.8}$$

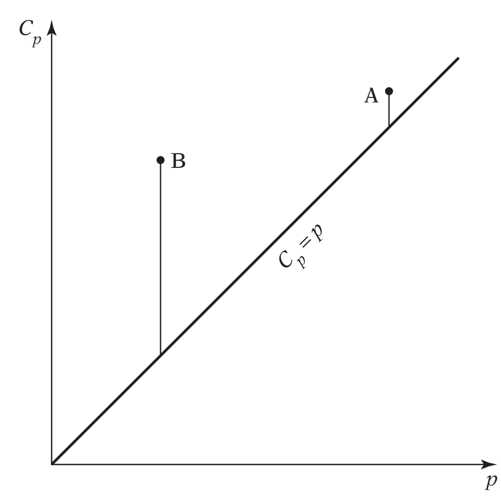
Al elegir un modelo de acuerdo con el criterio $C_p$, se debe buscar un modelo con un valor bajo de $C_p$, aproximadamente igual que $p$. En otras palabras, si seguimos el principio de parsimonia, elegiremos un modelo con $p$ regresoras ($p < k$) que proporcione un ajuste adecuado a los datos.

En la práctica, se suele graficar $C_p$ calculado de (13.9.7) en función de $p$. Un modelo "adecuado" se mostrará como un punto cercano a la línea $C_p = p$, como se observa en la figura 13.3, en la cual se ve que el modelo A es preferible al B, pues está más cerca a la línea $C_p = p$ que el B.

## Advertencia sobre los criterios de selección de modelos

Estudiamos varios criterios para seleccionar modelos. Pero estos criterios se deben considerar complementos de las diversas pruebas de especificación vistas en este capítulo. Algunos criterios analizados son meramente descriptivos y pueden carecer de propiedades teóricas firmes. Incluso a algunos se les puede imputar el cargo de recurrir a la minería de datos. Sin embargo, son tan comunes que el lector debe conocerlos. Ninguno de estos criterios es superior a los demás.[40] El

**FIGURA 13.3**
Gráfico de la $C_p$ de Mallows.




[39] Norman D. Draper y Harry Smith, *Applied Regression Analysis,* 3a. ed., John Wiley & Sons, Nueva York, 1998, p. 332. Consulte este libro para algunos ejemplos resueltos de $C_p$.

[40] Hay un análisis útil sobre el tema en Francis X. Diebold, *Elements of Forecasting,* 2a. ed., South Western, 2001, pp. 83-89. Respecto del balance, Diebold recomienda el criterio CIS.

software más moderno ahora incluye los criterios $R^2$, $R^2$ ajustada, CIA y CIS. El criterio de la $C_p$ de Mallows todavía no suele incluirse, aunque se obtiene con facilidad a partir de su definición.

### Pronóstico ji cuadrada ($\chi^2$)

Suponga que tenemos un modelo de regresión basado en $n$ observaciones y además deseamos pronosticar con él los valores (medios) de la regresada para $t$ observaciones adicionales. Siempre es aconsejable guardar parte de los datos muestrales para ver la forma en que el modelo estimado pronostica las observaciones no incluidas en la muestra, el periodo posmuestra:

Ahora el pronóstico $\chi^2$ se define como sigue:

$$\text{Pronóstico, } \chi^2 = \frac{\sum_{n+1}^{n+t} \hat{u}_i^2}{\hat{\sigma}^2} \tag{13.9.9}$$

donde $\hat{u}_i$ es el error de pronóstico para el periodo $i$ $(= n+1, n+2, \ldots, +n+t)$, con los parámetros obtenidos de la regresión ajustada y los valores de las regresoras en el periodo posmuestra. $\hat{\sigma}^2$ es el estimador usual de MCO para $\sigma^2$ basada en la regresión ajustada.

Si nuestra hipótesis es que los valores de los parámetros no cambiaron entre los periodos de la muestra y la posmuestra, podemos demostrar que el estadístico de (13.9.9) sigue la distribución ji cuadrada con $t$ grados de libertad, donde $t$ es el número de periodos para los que se realizó el pronóstico. Como señalan Charemza y Deadman, la prueba del pronóstico $\chi^2$ tiene un poder estadístico débil, lo cual significa que la probabilidad de que la prueba rechace correctamente una hipótesis nula falsa es baja y por tanto la prueba debe utilizarse más como indicador que como prueba definitiva.[41]

## 13.10 Otros temas relacionados con la creación de modelos econométricos

Como señalamos en la introducción de este capítulo, el tema de la construcción de modelos y de las pruebas de diagnóstico es tan amplio y complejo que hay libros especializados al respecto. En la sección anterior vimos en forma breve algunos temas importantes de esta área. En esta sección proseguimos con unos cuantos temas que a los investigadores les pueden parecer útiles en la práctica. En particular, estudiaremos los siguientes temas: 1) **valores atípicos, apalancamientos e influencia;** 2) **mínimos cuadrados recursivos** y 3) **prueba de falla de predicción de Chow**. Necesariamente, el análisis de cada uno será sucinto.

### Valores atípicos, apalancamiento e influencia[42]

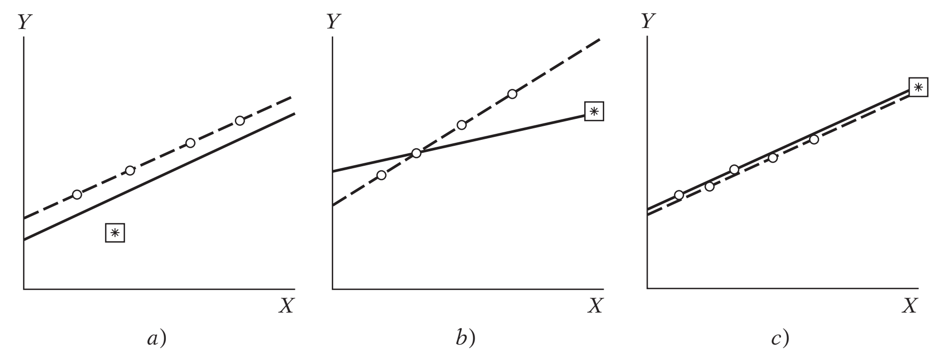
Recuerde que, al reducir la suma de cuadrado residual (SCR), los MCO dan igual ponderación a cada observación en la muestra. Pero cada una de éstas quizá no tenga igual efecto en los resultados de la regresión debido a la presencia de tres tipos de puntos de datos especiales, llamados **valores atípicos, puntos de apalancamiento** y **puntos de influencia**. Es importante saber lo que son y cómo influyen en el análisis de regresión.

En el contexto de la regresión, un **valor atípico** puede definirse como una observación con un "gran residuo". Recuerde que $\hat{u}_i = (Y_i - \hat{Y}_i)$; es decir, el residuo representa la diferencia (positiva o negativa) entre el valor real de la regresada y su valor estimado a partir del modelo de regresión.

[41] Wojciech W. Charemza y Derek F. Deadman, *New Directions in Econometric Practice: A General to Specific Modelling, Cointegration and Vector Autoregression,* 2a. ed., Edward Elgar, 1997, p. 30. Véase también pp. 250-252 para sus puntos de vista sobre diversos criterios en la selección de modelos.

[42] El siguiente análisis recibió la influencia de Chandan Mukherjee, Howard White y Marc Wyuts, *Econometrics and Data Analysis for Developing Countries,* Routledge, Nueva York, 1998, pp. 137-148.

**FIGURA 13.4** En cada inciso, las líneas continuas son las líneas de MCO para todos los datos, y las discontinuas son las líneas de MCO con el valor atípico, denotado por un ⊛, mismo que se omitió. En *a*), el valor atípico está cerca del valor medio de *X* y tiene un débil apalancamiento y poca influencia sobre los coeficientes de regresión. En *b*), el valor atípico está lejos del valor medio de *X* y tiene un fuerte apalancamiento, así como una influencia importante en los coeficientes de regresión. En *c*), el valor atípico tiene un gran apalancamiento pero poca influencia en los coeficientes de regresión debido a que está alineado al resto de las observaciones.




Fuente: Adaptado de John Fox, *op. cit*., p. 268.

Cuando decimos que un residuo es grande, lo comparamos con los demás residuos, y con mucha frecuencia ese residuo tan grande llama la atención de inmediato debido a su enorme distancia vertical respecto de la línea de regresión estimada. Observe que en el conjunto de datos puede haber más de un valor atípico. Ya vimos un ejemplo de lo anterior en el ejercicio 11.22, donde se pidió al lector hacer la regresión del cambio porcentual en los precios de acciones (*Y*) sobre el cambio porcentual en los precios al consumidor (*X*), con una muestra de 20 países. Hay que observar que Chile era un valor atípico.

Decimos que un dato ejerce **apalancamiento** (grande) si está desproporcionadamente distante de la mayor parte de los valores de una(s) regresora(s). ¿Por qué es importante un punto de apalancamiento? Porque es capaz de empujar la línea de regresión hacia él mismo, lo que distorsiona la pendiente de la línea de regresión. Si esto sucede, este punto (dato) se denomina de apalancamiento, un **punto de influencia**. La eliminación de tales puntos de datos de la muestra afecta de manera drástica a la línea de regresión. De vuelta al ejercicio 11.22, verá que si hace la regresión *Y* sobre *X*, incluso la observación para Chile, el coeficiente de la pendiente es positivo y "estadísticamente muy significativo". Pero si desecha la observación de Chile, el coeficiente de la pendiente es casi nulo. Por tanto, la observación sobre Chile tiene un apalancamiento y es una observación influyente.

Para aclarar aún más la naturaleza de los valores atípicos y los puntos de apalancamiento e influencia, observe el diagrama de la figura 13.4, el cual se explica por sí mismo.[43]

¿Cómo se trabaja con tales puntos de datos?, ¿sólo se debe eliminar y restringir la atención a los puntos de datos restantes? De acuerdo con Draper y Smith:

> El rechazo automático de los valores atípicos no siempre es sensato. A veces el valor atípico proporciona información que otros puntos de datos no suministran debido a que aquél surge de una combinación rara de circunstancias que puede revestir vital interés y requerir mayor investigación, en vez de rechazarlo. Por regla general, los valores atípicos deben rechazarse sólo si se originan por equívocos de registro, observaciones erróneas o un mal montaje de los aparatos [en un experimento físico]. De otro modo, se requiere una investigación cuidadosa.[44]

[43] Adaptado de John Fox, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage Publications, California, 1997, p. 268.

[44] Norman R. Draper y Harry Smith, *op. cit.,* p. 76.

¿Con qué pruebas se pueden detectar los valores atípicos y los puntos de apalancamiento? En la bibliografía hay varias, pero no las estudiaremos en este libro porque se desvían mucho del tema.[45] Los paquetes de software como SHAZAM y MICROFIT cuentan con rutinas para detectar los valores atípicos y los puntos de apalancamiento y de influencia.

## Mínimos cuadrados recursivos

En el capítulo 8 vimos la estabilidad estructural de un modelo de regresión que implicaba datos de series de tiempo, y mostramos la **prueba de Chow** que cumple con este propósito. De manera específica, estudiamos en ese capítulo una función de ahorro simple (el ahorro en función del ingreso) en Estados Unidos de 1970 a 2005. Vimos que la relación entre ahorro e ingreso tal vez cambió alrededor de 1982. Al conocer el punto crítico estructural, se confirmó mediante la prueba de Chow.

Pero, ¿qué sucede si no conocemos el punto de inflexión estructural? En este caso se utilizan los **mínimos cuadrados recursivos (MCR)**. La idea básica es muy sencilla: mediante la regresión de ahorro-ingreso.

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

donde $Y =$ ahorro y $X =$ ingreso en una muestra de 1970 a 2005. (Véase los datos de la tabla 8.11.)

Suponga que primero utilizamos los datos de 1970 a 1974 y estimamos la función ahorro, para obtener los estimados de $\beta_1$ y $\beta_2$. Luego utilizamos los datos de 1970 a 1975 y de nuevo estimamos la función ahorro para obtener los estimados de los dos parámetros. Más adelante empleamos los datos de 1970 a 1976 y volvemos a estimar el modelo de ahorro. Continuamos añadiendo así puntos de datos sobre $Y$ y $X$ hasta agotar la muestra. Como es de imaginarse, cada regresión proporciona un nuevo conjunto de estimaciones para $\beta_1$ y $\beta_2$. Si graficamos los valores estimados de estos parámetros respecto de cada iteración, vemos cómo cambian los parámetros estimados. Si el modelo en consideración es estructuralmente estable, las variaciones de los valores estimados de los dos parámetros serán mínimas y en esencia aleatorias. No obstante, si los valores estimados de los parámetros cambian en forma significativa, esto indica un rompimiento estructural. Por tanto, los MCR constituyen una herramienta útil con las series de tiempo, pues el tiempo está ordenado cronológicamente. También es una herramienta útil de diagnóstico en los datos transversales, donde los datos están ordenados por alguna variable de "tamaño" o "escala", como el empleo o el tamaño de los activos de una empresa. En el ejercicio 13.30 se pide al lector aplicar los MCR a los datos de ahorro de la tabla 8.11.

Los paquetes de software como SHAZAM, *EViews* y MICROFIT ya estiman en forma rutinaria los mínimos cuadrados recursivos. Asimismo, los MCR también generan **residuos recursivos**, en los que se basan diversas pruebas de diagnóstico.[46]

## Prueba de la falla de predicción de Chow

Ya analizamos en el capítulo 8 la estabilidad estructural de Chow. Demostró que esta prueba se puede modificar para comprobar el poder predictivo de un modelo de regresión. Veamos de nuevo la regresión de ahorro-ingreso en Estados Unidos de 1970 a 1995.

[45] Aquí se mencionan algunas fuentes: Alvin C. Rencher, *Linear Models in Statistics,* John Wiley & Sons, Nueva York, 2000, pp. 219-224; A.C. Atkinson, *Plots, Transformations and Regressions: An Introduction to Graphical Methods of Diagnostic Regression Analysis,* Oxford University Press, Nueva York, 1985, cap. 3; Ashis Sen y Muni Srivastava, *Regression Analysis: Theory, Methods, and Applications,* Springer-Verlag, Nueva York, 1990, cap. 8, y John Fox, *op. cit.,* cap. 11.

[46] Para más detalles, véase Jack Johnston y John DiNardo, *Econometric Methods,* 4a. ed., McGraw-Hill, Nueva York, 1997, pp. 117-121.

Suponga que estimamos la regresión ahorro-ingreso de 1970 a 1981 y obtenemos $\hat{\beta}_{1,70\text{-}81}$ y $\hat{\beta}_{2,70\text{-}81}$, las estimaciones para los coeficientes del intercepto y de la pendiente con base en los datos de 1970 a 1981. Ahora, con los verdaderos valores del ingreso de 1982 a 1995 y los valores del intercepto y la pendiente de 1970 a 1981, se predecimos los valores de ahorro de cada año de 1982 a 1995. La idea es que, si no hay un cambio estructural significativo en los valores de los parámetros, los valores de ahorro estimados de 1982 a 1995, con base en las estimaciones de los parámetros del periodo anterior, no deben ser muy distintos de los valores reales de ahorro que prevalecieron en el último periodo mencionado. Por supuesto, si hay una enorme diferencia entre los valores del ahorro pronosticados y los reales en el último periodo, surgirán dudas respecto de la estabilidad de la relación entre ahorro e ingresos para todo el periodo de los datos.

Mediante la prueba $F$ se demuestra si la diferencia entre el valor del ahorro real y estimado es pequeña o grande, de la siguiente forma:

$$F = \frac{\left(\sum \hat{u}_t^{*2} - \sum \hat{u}_t^2\right)/n_2}{\left(\sum \hat{u}_t^2\right)/(n_1 - k)} \tag{13.10.1}$$

donde $n_1$ = número de observaciones en el primer periodo (1970-1981), en las que se basa la regresión inicial, $n_2$ = número de observaciones en el segundo periodo pronosticado, $\sum \hat{u}_t^{*2}$ = SCR, cuando la ecuación se estima para todas las observaciones ($n_1 + n_2$), y $\sum \hat{u}_t^2$ = SCR cuando la ecuación se estima para las primeras $n_1$ observaciones y $k$ es el número de parámetros estimados (dos para este caso). Si los errores son independientes y están distribuidos de manera idéntica y normal, el estadístico $F$ dado en (13.10.1) sigue la distribución $F$, con $n_2$ y $n_1$ gl, respectivamente. En el ejercicio 13.31 se pide al lector aplicar la prueba de falla de predicción de Chow a fin de averiguar si la relación ahorro-ingreso en verdad cambió. A propósito, observe la similitud entre esta prueba y el pronóstico $\chi^2$ analizado antes.

## Datos faltantes

En el trabajo aplicado no es raro descubrir que a veces faltan observaciones de los datos de la muestra. Por ejemplo, en los datos de series de tiempo puede haber lagunas debido a circunstancias especiales. Durante la Segunda Guerra Mundial no hubo datos sobre algunas variables macroeconómicas o no se publicaron por razones estratégicas. En los datos transversales no es extraño descubrir que falta información sobre las variables de algunos individuos, en especial en los datos recopilados de encuestas por cuestionarios. En los datos de paneles, asimismo, algunos encuestados se retiran a la larga o no proporcionan información en todas las preguntas.

Sea cual fuere la razón, los datos faltantes son un problema que enfrenta todo investigador de vez en cuando. La pregunta es cómo tratar los datos faltantes. ¿Hay alguna forma de **asignar valores** a las observaciones faltantes?

No es fácil responder. Si bien existen algunas soluciones complicadas que se recomiendan en la bibliografía, no las estudiaremos aquí debido a su complejidad.[47] Sin embargo, analizaremos dos casos.[48] En el primero, las razones por las que hay datos faltantes son independientes de las observaciones disponibles, situación que Darnell denomina "caso ignorable". En el segundo caso, no sólo los datos disponibles están incompletos, sino que las observaciones faltantes se relacionan de manera sistemática con los datos disponibles. Este caso es más grave, pero puede ser resultado del *sesgo de autoselección*, es decir, los datos observados no se recopilan de manera en verdad aleatoria.

[47] Para un tratamiento minucioso y avanzado del tema, véase A. Colin Cameron y Pravin K. Trivedi, *Microeconometrics: Methods and Applications,* Cambridge University Press, Nueva York, 2005, capítulo 27, pp. 923-941.

[48] El siguiente análisis se basa en Adrian C. Darnell, *A Dictionary of Econometrics,* Edward Elgar Publishing, Lyne, Reino Unido, 1994, pp. 256-258.

En el caso ignorable se pueden simplemente pasar por alto las observaciones faltantes y usar las disponibles. La mayoría de los paquetes de software estadístico lo hace de forma automática. Por supuesto, en este caso el tamaño de la muestra se reduce y es posible que no se obtengan estimaciones precisas de los coeficientes de regresión. Sin embargo, con los datos disponibles pueden aclararse las observaciones faltantes. Aquí veremos tres posibilidades.

1. De un número total de $N$ observaciones tenemos datos completos sobre $N_1$ ($N_1 < N$) tanto para la variable regresada como para $k$ regresoras denotadas por $\boldsymbol{Y_1}$ y $\boldsymbol{X_1}$, respectivamente. ($\boldsymbol{Y_1}$ es un vector de $N_1$ observaciones, y $\boldsymbol{X_1}$, un vector renglón de $k$ regresoras).
2. En algunas observaciones ($N_2 < N$) existen datos completos para la regresada, denotada por $\boldsymbol{Y_2}$, pero observaciones incompletas en algunas $\boldsymbol{X_2}$ (de nuevo, se trata de vectores).
3. En algunas observaciones ($N_3 < N$) no hay datos sobre $Y$, pero tenemos datos completos sobre $X$, denotados por $\boldsymbol{X_3}$.

En el primer caso, la regresión de $\boldsymbol{Y_1}$ sobre $\boldsymbol{X_1}$ produce estimaciones de los coeficientes de regresión insesgados, pero tal vez no sean eficientes porque ignoramos $N_2$ y $N_3$ observaciones. Los otros dos casos son muy complicados y corresponde al lector consultar las referencias.[49]

## 13.11 Ejemplos para concluir

Terminamos con dos ejemplos que ilustran uno o más de los puntos planteados. El primer ejemplo, sobre determinación de salarios, usa datos transversales, y el segundo, que considera la función de consumo real de Estados Unidos, datos de series de tiempo.

### 1. Un modelo de determinación de salarios por hora

Para examinar los factores que determinan los salarios por hora consideraremos un modelo salarial tipo Mincer, popular ya entre los economistas especializados en asuntos laborales. Este modelo adopta la siguiente forma:[50]

$$\ln \text{salario}_i = \beta_1 + \beta_2 \text{Esc}_i + \beta_3 \text{Exp}_i + \beta_4 \text{Fe}_i + \beta_5 \text{NB}_i + \beta_6 \text{Sind}_i + \beta_7 \text{Sem}_i + u_i \tag{13.11.1}$$

donde ln salario = logaritmo natural del salario por hora ($), Esc = escolaridad en años, Exp = experiencia en el mercado laboral, Fe = 1 si es femenino, 0 en otro caso, NB = 1 si el trabajador no es blanco, 0 en otro caso, Sind = 1 si es trabajador sindicalizado, 0 en otro caso, y Sem = 1 si es trabajador que no recibe salario por hora, 0 en otro caso. Para los trabajadores que no reciben salario por hora, éste se calcula como el salario semanal dividido entre el número acostumbrado de horas trabajadas.

Se pueden añadir muchas variables más a este modelo, como origen étnico, estado civil, número de hijos menores de 6 años y riqueza o ingreso no procedente del trabajo. Por el momento trabajaremos con el modelo de la ecuación (13.11.1).

Los datos corresponden a 1 289 personas entrevistadas en marzo de 1985 como parte de la *Current Population Survey* (*CPS*), que periódicamente realiza la Oficina del Censo de Estados Unidos. Paul Rudd recopiló estos datos originalmente.[51]

---

[49] Además de las referencias ya citadas, véase A.A. Afifi y R.M. Elashoff, "Missing Observations in Multivariate Statistics", *Journal of the American Statistical Association,* vol. 61, 1966, pp. 595-604, y vol. 62, 1967, pp. 10-29.

[50] Véase J. Mincer, *School, Experience and Earnings,* Columbia University Press, Nueva York, 1974.

[51] Paul A. Rudd, *An Introduction to Classical Econometric Theory,* Oxford University Press, Nueva York, 2000. No se incluyeron datos sobre edad porque son muy colineales con la experiencia laboral.

A priori, esperaríamos que la escolaridad y la experiencia tuviesen efecto positivo en los salarios. Se espera que las variables dicótomas Fe y NB tengan efecto negativo en los salarios si existe algún tipo de discriminación, y que Sind tenga efecto positivo debido a la incertidumbre del ingreso.

Cuando todas las variables dicótomas toman un valor de cero, la ecuación (13.11.1) se reduce a

$$\ln \text{salario}_i = \beta_1 + \beta_2 \text{Esc}_i + \beta_3 \text{Exp}_i + u_i \tag{13.11.2}$$

que es la función de salario de un trabajador blanco, masculino, no sindicalizado y que percibe salario por hora. Esta es la categoría base, o *de referencia*.

A continuación presentamos los resultados de la regresión y luego los analizamos.

**TABLA 13.4 Resultados de la regresión de *EViews* con base en la ecuación (13.11.1)**

Variable dependiente: LS
Método: Mínimos cuadrados
Muestra: 1-1, 289
Observaciones incluidas: 1 289

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | 1.037880 | 0.074370 | 13.95563 | 0.0000 |
| Esc | 0.084037 | 0.005110 | 16.44509 | 0.0000 |
| Exp | 0.011152 | 0.001163 | 9.591954 | 0.0000 |
| Fe | -0.234934 | 0.026071 | -9.011170 | 0.0000 |
| NB | -0.124447 | 0.036340 | -3.424498 | 0.0006 |
| Sind | 0.207508 | 0.036265 | 5.721963 | 0.0000 |
| Sem | 0.228725 | 0.028939 | 7.903647 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.376053 | Media de la variable dependiente | 2.342416 |
| *R* cuadrada ajustada | 0.373133 | Desviación estándar de la variable dependiente | 0.586356 |
| Error estándar de la regresión | 0.464247 | Criterio de información de Akaike | 1.308614 |
| Suma de cuadrados residual | 276.3030 | Criterio de Schwarz | 1.336645 |
| Log verosimilitud | -836.4018 | Criterio de Hannan-Quinn | 1.319136 |
| Estadístico *F* | 128.7771 | Estad. Durbin-Watson | 1.977004 |
| Prob. (estadístico *F*) | 0.000000 | | |

Lo primero que se observa es que todos los coeficientes estimados son muy significativos en lo *individual*, pues los valores *p* son muy bajos. El valor *F* también es muy alto, lo que indica que también, en conjunto, todas las variables son estadísticamente importantes.

En comparación con el trabajador de referencia, el salario promedio de una trabajadora y de un trabajador no blanco es inferior. Los trabajadores sindicalizados y los que perciben salario semanal, en promedio, ganan más.

¿Es adecuado el modelo (13.11.1) dadas las variables consideradas? ¿Es posible que las trabajadoras no blancas ganen menos que los trabajadores blancos? ¿Es posible que las trabajadoras no blancas y no sindicalizadas ganen menos que las trabajadoras blancas sindicalizadas? En otras palabras, ¿hay efectos de interacción entre las regresoras cuantitativas y las variables dicótomas?

Los paquetes estadísticos responden a estas preguntas. Por ejemplo, *EViews* cuenta con esta capacidad. Después de estimar un modelo, si uno cree que se le pueden agregar algunas variables pero no está seguro de su importancia, puede ejecutar la prueba de *variables omitidas*.

Para demostrar esto, suponga que estimamos la ecuación (13.11.1) y deseamos averiguar si los productos de Fe y NB, Fe y Sind, y Fe y Sem deben agregarse al modelo para tomar en cuenta la interacción entre las variables explicativas. Con la rutina de *EViews6* obtenemos la siguiente

respuesta: la *hipótesis nula* es que estas tres variables añadidas no tienen efecto en el modelo estimado.

Como es de suponer, la prueba $F$ (estudiada en el capítulo 8) sirve para evaluar la contribución marginal, o incremental, de las variables añadidas y probar la hipótesis nula. En este ejemplo, los resultados son los siguientes:

**TABLA 13.5** **Resultados parciales de *EViews* mediante interacciones**

| Variables omitidas: Fe*NB Fe*Sind Fe*Sem | | | |
|---|---|---|---|
| Estadístico $F$ | 0.805344 | Prob. $F$ (3 1279) | 0.4909 |
| Log razón de verosimilitud | 2.432625 | Prob. ji cuadrada (3) | 0.4876 |

No rechazamos la hipótesis nula de que la interacción entre mujeres y no blancos, mujeres y sindicalizados y mujeres y trabajadores que perciben salario semanal, colectivamente, no tiene efecto significativo en el modelo estimado que se presenta en la tabla 13.4, pues el valor estimado de $F$ de 0.8053 no es estadísticamente significativo y el valor $p$ es de aproximadamente 49 por ciento.

Queda al lector la tarea de probar otras combinaciones de regresoras para evaluar su contribución al modelo semanal.

Antes de proseguir, el modelo (13.11.1) indica que la influencia de la experiencia sobre el logaritmo de los salarios es lineal, es decir, si las demás variables se mantienen constantes, el incremento relativo en los salarios (recuerde que la regresada está en forma logarítmica) es igual por cada año de aumento de experiencia laboral. Este supuesto puede ser verdadero a través de varios años de experiencia, pero como indica la economía laboral básica, conforme los trabajadores envejecen, el aumento de la tarifa salarial se reduce. Para comprobar si así sucede en este ejemplo, agregamos el término de experiencia al cuadrado al modelo inicial, con los siguientes resultados:

**TABLA 13.6** **Resultados de *EViews* con experiencia al cuadrado**

Variable dependiente: LS
Método: Mínimos cuadrados
Muestra: 1-1, 289
Observaciones incluidas: 1 289

| | Coeficiente | Error estándar | Estadístico $t$ | Prob. |
|---|---|---|---|---|
| C | 0.912279 | 0.075151 | 12.13922 | 0.0000 |
| Esc | 0.079867 | 0.005051 | 15.81218 | 0.0000 |
| Exp | 0.036659 | 0.003800 | 9.647230 | 0.0000 |
| Fe | -0.228848 | 0.025606 | -8.937218 | 0.0000 |
| NB | -0.121805 | 0.035673 | -3.414458 | 0.0007 |
| Sind | 0.199957 | 0.035614 | 5.614579 | 0.0000 |
| Sem | 0.222549 | 0.028420 | 7.830675 | 0.0000 |
| Exp*Exp | -0.000611 | 8.68E-05 | -7.037304 | 0.0000 |

| | | | |
|---|---|---|---|
| $R$ cuadrada | 0.399277 | Media de la variable dependiente | 2.342416 |
| $R$ cuadrada ajustada | 0.395995 | Desviación estándar de la variable dependiente | 0.586356 |
| Error estándar de la regresión | 0.455703 | Criterio de información de Akaike | 1.272234 |
| Suma de cuadrados residual | 266.0186 | Criterio de Schwarz | 1.304269 |
| Log verosimilitud | -811.9549 | Criterio de Hannan-Quinn | 1.284259 |
| Estadístico $F$ | 121.6331 | Estad. Durbin-Watson | 1.971753 |
| Prob. (estadístico $F$) | 0.000000 | | |

El término de experiencia al cuadrado no sólo es negativo, sino también muy significativo estadísticamente. Además, concuerda con el comportamiento del mercado laboral; con el tiempo, la tasa de crecimiento de los salarios se desacelera $\left(\frac{\partial lw}{\partial \text{EXP}} = 0.0366 - 0.0012\text{EXP}\right)$.

Aquí se abre la oportunidad para explicar los criterios de Akaike y Schwartz. Al igual que $R^2$, son pruebas de la bondad del ajuste del modelo estimado; la diferencia radica en que según el criterio $R^2$, cuanto mayor sea el valor, el modelo explicará mejor el comportamiento de la variable regresada. Por otra parte, según los criterios de Akaike y Schwartz, cuanto más bajo sea el valor de estos estadísticos, mejor será el modelo.

Por supuesto, todos estos criterios son significativos si deseamos comparar dos o más modelos. Así, si comparamos el modelo de la tabla 13.4 con el de la tabla 13.6, que tiene la experiencia elevada al cuadrado como regresora adicional, se observa que el modelo de la tabla 13.6 es preferible al de la tabla 13.4 con base en los tres criterios.

A propósito, note que en los dos modelos los valores de $R^2$ parecen "bajos", pero estos valores bajos se observan por lo general en datos transversales con un gran número de observaciones. Sin embargo, tenga en cuenta que este valor "bajo" de $R^2$ es estadísticamente significativo, pues, en los dos modelos, el estadístico $F$ calculado es muy significativo (recuerde la relación entre $F$ y $R^2$ del capítulo 8).

Continuemos con el modelo ampliado de la tabla 13.6. Aunque el modelo parece satisfactorio, exploraremos dos cuestiones. Primera, en vista de que se trata de datos transversales, hay muchas probabilidades de que el modelo sufra de heteroscedasticidad. En consecuencia, es preciso averiguar si es así. Aplicamos varias de las pruebas de heteroscedasticidad estudiadas en el capítulo 11 y comprobamos que el modelo, en efecto, tiene heteroscedasticidad. El lector debe corroborar esta afirmación.

Para corregir esta heteroscedasticidad obtenemos *los errores estándar consistentes con heteroscedasticidad de White*, que examinamos en el capítulo 11. Los resultados se presentan en la siguiente tabla.

**TABLA 13.7** **Resultados de *EViews* mediante errores estándar corregidos de White**

Variable dependiente: LS
Método: Mínimos cuadrados
Muestra: 1-1, 289
Observaciones incluidas: 1 289
Errores estándar consistentes con heteroscedasticidad de White y covarianza

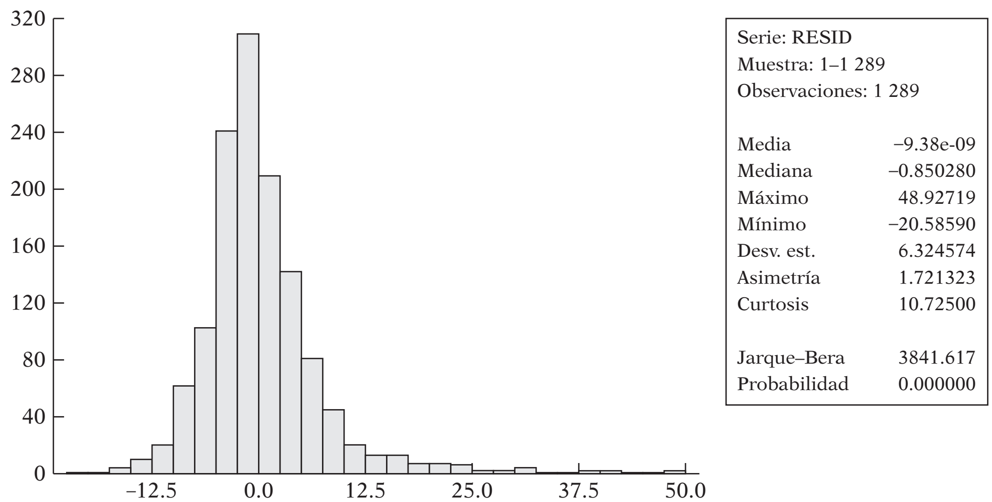
| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | 0.912279 | 0.077524 | 11.76777 | 0.0000 |
| Esc | 0.079867 | 0.005640 | 14.15988 | 0.0000 |
| Exp | 0.036659 | 0.003789 | 9.675724 | 0.0000 |
| Fe | -0.228848 | 0.025764 | -8.882625 | 0.0000 |
| NB | -0.121805 | 0.033698 | -3.614573 | 0.0003 |
| Sind | 0.199957 | 0.029985 | 6.668458 | 0.0000 |
| Sem | 0.222549 | 0.031301 | 7.110051 | 0.0000 |
| Exp*Exp | -0.000611 | 9.44E-05 | -6.470218 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.399277 | Media de la variable dependiente | 2.342416 |
| *R* cuadrada ajustada | 0.395995 | Desviación estándar de la variable dependiente | 0.586356 |
| Error estándar de la regresión | 0.455703 | | |
| Suma de cuadrados residual | 266.0186 | Criterio de información de Akaike | 1.272234 |
| Log verosimilitud | -811.9549 | Criterio de Schwarz | 1.304269 |
| Estadístico *F* | 121.6331 | Criterio de Hannan-Quinn | 1.284259 |
| Prob. (estadístico *F*) | 0.000000 | Estad. Durbin-Watson | 1.971753 |

Como es de esperar, se advierten algunos cambios en los errores estándar estimados, aunque esto no modifica la conclusión de que todas las regresoras son importantes, tanto en lo individual como en lo colectivo, para explicar el comportamiento de los salarios relativos.

A continuación examinaremos si los términos de error están distribuidos normalmente. En la figura 13.5 se presenta el histograma de los residuos obtenidos del modelo de la tabla 13.7. El

**FIGURA 13.5**
Histograma de los residuos obtenidos de la regresión de la tabla 13.7.




estadístico de Jarque-Bera rechaza la hipótesis de que los errores están distribuidos normalmente, pues el estadístico de JB es alto y el valor *p* es prácticamente cero: observe que para una variable distribuida normalmente, los coeficientes de asimetría y curtosis son 0 y 3, respectivamente.

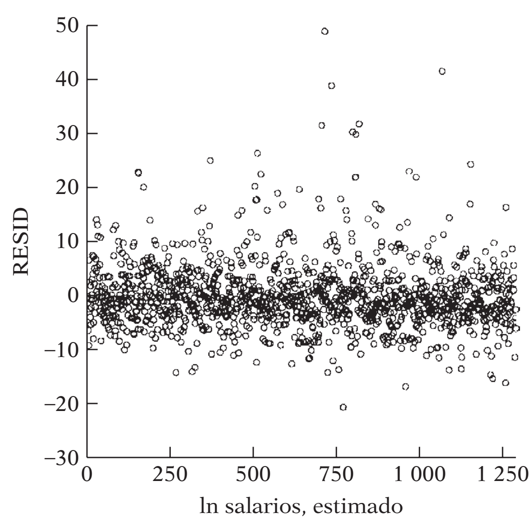
¿Qué sigue? El procedimiento de pruebas de hipótesis se ha basado hasta el momento en el supuesto de que el término de perturbación, o error, en el modelo de regresión está distribuido normalmente. ¿Significa esto que no se pueden usar legítimamente las pruebas *t* y *F* para probar las hipótesis de la regresión de los salarios?

La respuesta es **no**. Como señalamos en el capítulo, los estimadores de MCO tienen distribución normal asintótica, con la advertencia apuntada en el capítulo, es decir, el término de error tiene varianza finita, es homoscedástico y el valor medio del término de error, dados los valores de las variables explicativas, es cero. Como resultado, podemos seguir usando las pruebas *t* y *F* acostumbradas en tanto la muestra sea razonablemente grande. A propósito, debe señalarse que no necesitamos el supuesto de normalidad para obtener los estimadores de MCO. Aun sin el supuesto de normalidad, los estimadores de MCO son los mejores estimadores lineales insesgados (MELI) según los supuestos de Gauss-Markov.

¿De qué tamaño es una muestra grande? No hay respuesta definitiva, pero el tamaño de la muestra de 1 289 observaciones de la regresión de los salarios parece razonablemente grande.

¿Hay "valores atípicos" en la regresión de los salarios? Es posible darse una idea con el gráfico de la figura 13.6, que presenta los valores reales y estimados de la variable dependiente

**FIGURA 13.6**
Residuos y valores estimados de la variable dependiente, ln salarios.



(ln salarios) y los residuos, que son las diferencias entre los valores reales y los estimados de la regresada.

Aunque el valor medio de los residuos es siempre cero (¿por qué?), la gráfica de la figura 13.6 muestra varios residuos que dan la impresión de ser grandes (en valor absoluto) en comparación con la mayoría de ellos. Es posible que haya valores atípicos en los datos. Se proporcionan las estadísticas básicas de las tres variables cuantitativas en la tabla 13.8 para que el lector decida si, en efecto, hay valores atípicos.

**TABLA 13.8**

Muestra: 1-1, 289

| | Sal | Esc | Exp |
|---|---|---|---|
| Media | 12.36585 | 13.14507 | 18.78976 |
| Mediana | 10.08000 | 12.00000 | 18.00000 |
| Máximo | 64.08000 | 20.00000 | 56.00000 |
| Mínimo | 0.840000 | 0.000000 | 0.000000 |
| Desv. est. | 7.896350 | 2.813823 | 11.66284 |
| Asimetría | 1.848114 | -0.290381 | 0.375669 |
| Curtosis | 7.836565 | 5.977464 | 2.327946 |
| Jarque-Bera | 1990.134 | 494.2552 | 54.57664 |
| Probabilidad | 0.000000 | 0.000000 | 0.000000 |
| Suma | 15 939.58 | 16 944.00 | 24 220.00 |
| Suma desv. al cuadrado | 80 309.82 | 10 197.87 | 175 196.0 |
| Observaciones | 1 289 | 1 289 | 1 289 |

## 2. Función de consumo real de Estados Unidos, 1947-2000

En el capítulo 10 consideramos la función de consumo en Estados Unidos de 1947 a 2000. La forma específica de la función de consumo fue:

$$\ln \text{GT}_t = \beta_1 + \beta_2 \ln \text{ID}_t + \beta 3 \ln \text{Riq} + \beta 4\text{Interés}_t + u_t \qquad \textbf{(13.11.3)}$$

donde GT, ID, Riqueza e Interés son, respectivamente, gasto total de consumo, ingreso personal disponible, riqueza y tasa de interés, todos en términos reales. Los resultados basados en los datos son los siguientes:

**TABLA 13.9** **Resultados de la ecuación de regresión (13.11.3)**

Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | -0.467711 | 0.042778 | -10.93343 | 0.0000 |
| LOG (ID) | 0.804873 | 0.017498 | 45.99836 | 0.0000 |
| LOG (RIQUEZA) | 0.201270 | 0.017593 | 11.44060 | 0.0000 |
| INTERÉS | -0.002689 | 0.000762 | -3.529265 | 0.0009 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.999560 | Media de la variable dependiente | 7.826093 |
| *R* cuadrada ajustada | 0.999533 | Desviación estándar de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.011934 | Criterio de información de Akaike | -5.947703 |
| Suma de cuadrados residual | 0.007121 | Criterio de Schwarz | -5.800371 |
| Log verosimilitud | 164.5880 | Criterio de Hannan-Quinn | -5.890883 |
| Estadístico *F* | 37 832.59 | Estadístico de Durbin-Watson | 1.289219 |
| Prob. (estadístico *F*) | 0.000000 | | |

Como GT, ID y Riqueza entran en forma de logaritmo, los coeficientes estimados de la pendiente de ID y Riqueza son, respectivamente, las elasticidades del ingreso y la riqueza. Como es de esperar, estas elasticidades son positivas y muy significativas estadísticamente. En términos numéricos, las elasticidades del ingreso y de la riqueza son casi 0.80 y 0.20. El coeficiente de la variable tasa de interés representa semielasticidad (¿por qué?). Si las demás variables permanecen constantes, los resultados demuestran que si la tasa de interés aumenta un punto porcentual, en promedio, el gasto de consumo real se reduce alrededor de 0.27%. Observe que la semielasticidad estimada también es muy significativa en términos estadísticos.

Aprecie algunos estadísticos del resumen. El valor de $R^2$ es muy alto y llega casi a 100%. El valor $F$ también es muy significativo estadísticamente, lo que indica que no sólo en lo individual, sino también en conjunto, todas las variables explicativas tienen efecto significativo sobre el gasto de consumo.

No obstante, el estadístico de Durbin-Watson indica que los errores del modelo están serialmente correlacionados. Si consulta las tablas Durbin-Watson (tabla D.5, apéndice D), advertirá que para 55 observaciones (el número más cercano a 54) y tres variables explicativas, los valores críticos $d$ inferior y superior a 5% son 1.452 y 1.681. Como la $d$ observada en el ejemplo, 1.2892, se sitúa por debajo del valor crítico $d$ inferior, podemos concluir que los errores de esta función de consumo tienen correlación positiva. Este resultado no debe sorprender, pues en la mayoría de las regresiones de series de tiempo existe autocorrelación.

Pero antes de aceptar esta conclusión, veamos si existen errores de especificación. Como sabemos, en ocasiones la autocorrelación puede ser aparente porque se omitieron algunas variables importantes. Para ver si es así, consideremos la regresión obtenida en la tabla 13.10.

**TABLA 13.10**

Variable dependiente: LGT
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54

| | Coeficiente | Error estándar | Estadístico $t$ | Probabilidad |
|---|---|---|---|---|
| C | 2.689644 | 0.566034 | 4.751737 | 0.0000 |
| LID | 0.512836 | 0.054056 | 9.487076 | 0.0000 |
| LRiqueza | -0.205281 | 0.074068 | -2.771510 | 0.0079 |
| INTERÉS | -0.001162 | 0.000661 | -1.759143 | 0.0848 |
| LID*LRiqueza | 0.039901 | 0.007141 | 5.587986 | 0.0000 |

| | | | |
|---|---|---|---|
| $R$ cuadrada | 0.999731 | Media de la variable dependiente | 7.8260933 |
| $R$ cuadrada ajustada | 0.999709 | Desviación estándar de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.009421 | | |
| Suma de cuadrados residual | 0.004349 | Criterio de información de Akaike | -6.403689 |
| Log verosimilitud | 177.8996 | Criterio de Schwarz | -6.219524 |
| Estadístico $F$ | 45 534.94 | Criterio de Hannan-Quinn | -5.890883 |
| Probabilidad (estadístico $F$) | 0.000000 | Estadístico de Durbin-Watson | 1.530268 |

La variable adicional de este modelo es la interacción de los logaritmos del ingreso disponible y de la riqueza. El término de interacción es muy significativo. Observe que ahora la variable interés es menos significativa (el valor $p$ es de casi 8%), aunque conserva el signo negativo. Pero ahora el valor $d$ de Durbin Watson aumentó de casi 1.28 a alrededor de 1.53.

Los valores críticos $d$ a 5% son ahora 1.378 y 1.721. El valor $d$ observado de 1.53 se sitúa entre estos valores, lo que indica que, con base en el estadístico de Durbin-Watson, no podemos determinar si hay autocorrelación. Sin embargo, el valor $d$ observado está más cerca del límite

superior del valor *d*. Como señalamos en el capítulo sobre autocorrelación, algunos autores proponen usar el límite superior del estadístico *d* como el verdadero límite de significancia aproximado; por tanto, si el valor *d* calculado se sitúa por debajo del límite superior, hay indicios de autocorrelación positiva. Con base en ese criterio, en el presente caso podemos concluir que el modelo sufre de autocorrelación positiva.

También aplicamos la prueba de autocorrelación de Breusch-Godfrey que explicamos en el capítulo 12. Al sumar los dos términos rezagados de los residuos estimados en la ecuación (12.6.15) al modelo de la tabla 13.9, obtuvimos los siguientes resultados:

**TABLA 13.11**

Prueba ML de correlación serial de Breusch-Godfrey

| | | | |
|---|---|---|---|
| Estadístico *F* | 3.254131 | Prob. *F*(2, 48) | 0.0473 |
| *R* cuadrada obs.* | 6.447576 | Prob. ji cuadrada (2) | 0.0398 |

Variable dependiente: RESID
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54
Valores de los residuos rezagados faltantes de preprueba fijados en cero.

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | -0.006514 | 0.041528 | -0.156851 | 0.8760 |
| LID | -0.004197 | 0.017158 | -0.244619 | 0.8078 |
| LRiqueza | 0.004191 | 0.017271 | 0.242674 | 0.8093 |
| INTERÉS | 0.000116 | 0.000736 | 0.156964 | 0.8759 |
| RESID (-1) | 0.385190 | 0.151581 | 2.541147 | 0.0143 |
| RESID (-2) | -0.165609 | 0.154695 | -1.070556 | 0.2897 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.119400 | Media de la variable dependiente | -9.02E-17 |
| *R* cuadrada ajustada | 0.027670 | Desviación estándar de la variable dependiente | 0.011591 |
| Error estándar de la regresión | 0.011430 | Criterio de información de Akaike | -6.000781 |
| Suma de cuadrados residual | 0.006271 | Criterio de Schwarz | -5.779782 |
| Log verosimilitud | 168.0211 | Criterio de Hannan-Quinn | -5.915550 |
| Estadístico *F* | 1.301653 | Estadístico de Durbin-Watson | 1.848014 |
| Probabilidad (estadístico *F*) | 0.279040 | | |

El valor *F* reportado al principio de la tabla prueba la hipótesis de que los dos residuos rezagados incluidos en el modelo tienen valores cero. Esta hipótesis se rechaza porque el valor *F* es significativo en el nivel aproximado de 5 por ciento.

Para resumir, parece que hay autocorrelación en el término de error. Se pueden aplicar uno o más de los procedimientos analizados en el capítulo 12 para eliminar la autocorrelación. Sin embargo, para ahorrar espacio, se deja esa tarea al lector.

En la tabla 13.12 presentamos los resultados del análisis de regresión de los errores estándar CHA o Newey-West que toman en cuenta la autocorrelación. El tamaño de la muestra de 54 observaciones es lo bastante grande para usar los errores estándar CHA.

Si compara estos resultados con los que se presentan en la tabla 13.9, observará que los coeficientes de regresión siguen siendo los mismos, pero los errores estándar son un poco diferentes.

En este capítulo hablamos de la prueba de falla de predicción de Chow. Tenemos un periodo de muestra de 1947 a 2000. Durante este periodo se registraron varios ciclos económicos, en su mayoría breves. Por ejemplo, hubo una recesión en 1990 y otra en 2000. ¿El comportamiento del

## TABLA 13.12

Variable dependiente: LGT
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 54
Errores estándar CHA Newey-West y covarianza (truncamiento de rezago = 3)

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | -0.467714 | 0.043937 | -10.64516 | 0.0000 |
| LID | 0.804871 | 0.017117 | 47.02132 | 0.0000 |
| LRiqueza | 0.201272 | 0.015447 | 13.02988 | 0.0000 |
| INTERÉS | -0.002689 | 0.000880 | -3.056306 | 0.0036 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.999560 | Media de la variable dependiente | 7.826093 |
| *R* cuadrada ajustada | 0.999533 | Desviación estándar de la variable dependiente | 0.552368 |
| Error estándar de la regresión | 0.011934 | Criterio de información de Akaike | -5.947707 |
| Suma de cuadrados residual | 0.007121 | Criterio de Schwarz | -5.800374 |
| Log verosimilitud | 164.5881 | Criterio de Hannan-Quinn | -5.890886 |
| Estadístico *F* | 37 832.71 | Estadístico de Durbin-Watson | 1.289237 |
| Probabilidad (estadístico *F*) | 0.000000 | | |

gasto de consumo en relación con el ingreso, riqueza y la tasa de interés es distinto durante las recesiones?

Para aclarar esta pregunta, consideremos la recesión de 1990 y apliquemos la prueba de falla de predicción de Chow. Los detalles de esta prueba ya se analizaron en el capítulo. Con la prueba de falla de predicción de Chow en *EViews*, versión 6, obtenemos los resultados de la tabla 13.13.

## TABLA 13.13 **Prueba de falla de predicción de Chow**

Prueba de pronóstico de Chow: Pronóstico de 1991 a 2000

| | | | |
|---|---|---|---|
| Estadístico *F* | 1.957745 | Prob. *F* (10, 40) | 0.0652 |
| Log razón de verosimilitud | 21.51348 | Prob. ji cuadrada (10) | 0.0178 |

Variable dependiente: LGT
Método: Mínimos cuadrados
Muestra: 1947-2000
Observaciones incluidas: 44

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| C | -0.287952 | 0.095089 | -3.028236 | 0.0043 |
| LID | 0.853172 | 0.028473 | 29.96474 | 0.0000 |
| LRiqueza | 0.141513 | 0.033085 | 4.277239 | 0.0001 |
| INTERÉS | -0.002060 | 0.000804 | -2.562790 | 0.0143 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.999496 | Media de la variable dependiente | 7.659729 |
| *R* cuadrada ajustada | 0.999458 | Desviación estándar de la variable dependiente | 0.469580 |
| Error estándar de la regresión | 0.010933 | Criterio de información de Akaike | -6.107640 |
| Suma de cuadrados residual | 0.004781 | Criterio de Schwarz | -5.945441 |
| Log verosimilitud | 138.3681 | Criterio de Hannan-Quinn | -6.047489 |
| Estadístico *F* | 26 430.49 | Estadístico de Durbin-Watson | 1.262748 |
| Probabilidad (estadístico *F*) | 0.000000 | | |

El estadístico $F$ dado en la parte superior de la tabla 13.13 indica que probablemente no hay diferencia considerable en la función de consumo antes y después de 1990, pues el valor $p$ no es significativo en el nivel de 5%. Sin embargo, si elegimos el nivel de significancia de 10%, el valor $F$ es significativo estadísticamente.

Podemos examinar este problema de otra manera. En el capítulo 8 analizamos una prueba de estabilidad de parámetros. Para ver si hubo algún cambio estadísticamente significativo en los coeficientes de regresión de la función de consumo, aplicamos la prueba de Chow estudiada en la sección 8.7 del capítulo 8 y obtuvimos los resultados de la tabla 13.14.

**TABLA 13.14**
**Prueba de estabilidad paramétrica de Chow**

Prueba de puntos de ruptura estructural de Chow: 1990
Hipótesis nula: No hay rupturas en los puntos especificados
Regresoras que varían: Todas las variables de la ecuación
Muestra de la ecuación: 1947-200

| | | | |
|---|---|---|---|
| Estadístico *F* | 4.254054 | Prob. *F*(4, 46) | 0.0052 |
| Log razón de verosimilitud | 16.99654 | Prob. ji cuadrada (4) | 0.0019 |
| Estadístico de Wald | 17.01622 | Prob. ji cuadrada (4) | 0.0019 |

En apariencia, las funciones de consumo antes y después de 1990 son estadísticamente diferentes, pues el estadístico $F$ calculado, según la ecuación (8.7.4), es muy significativo estadísticamente, porque el valor $p$ es de sólo 0.0052.

Se invita al lector a aplicar las pruebas de estabilidad paramétrica y falla de predicción de Chow para determinar si varió la función de consumo antes y después de 2000. Para ello será necesario ampliar los datos más allá de 2000. Asimismo, observe que, para aplicar estas pruebas, el número de observaciones debe ser mayor que el número de coeficientes estimados.

Agotamos todas las pruebas de diagnóstico que se pueden aplicar a los datos de consumo. Sin embargo, el análisis hasta el momento debe dar una idea muy buena de cómo aplicar las diversas pruebas.

## 13.12 Errores no normales y regresoras estocásticas

En esta sección analizamos dos temas más o menos avanzados, a saber, la distribución no normal del término de error y las regresoras estocásticas, o aleatorias, y su importancia práctica.

### 1. ¿Qué pasa si el término de error no está distribuido normalmente?

En el *modelo clásico de regresión lineal normal* (MCRLN) explicado en el capítulo 4, supusimos que el término de error u sigue la distribución normal. Invocamos el **teorema central del límite (TCL)** para justificar el supuesto de normalidad. Debido a este supuesto, podemos establecer que los estimadores de MCO también están distribuidos normalmente. Como resultado, para probar hipótesis aplicamos las pruebas $t$ y $F$ sin importar el tamaño de la muestra. También explicamos cómo usar las pruebas de normalidad de **Jarque-Bera** y **Anderson-Darling** para averiguar si los errores estimados están distribuidos normalmente en cualquier aplicación práctica.

¿Qué sucede si los errores *no* están distribuidos normalmente? Podemos afirmar que los estimadores de MCO siguen siendo MELI, es decir, que son insesgados y en la clase de estimadores lineales exhiben varianza mínima. En principio, esto no debe sorprender, pues, para establecer el teorema de Gauss-Markov (MELI), no necesitamos el supuesto de normalidad.

Entonces, ¿cuál es el problema?

El problema es que se necesitan las **distribuciones de muestreo**, o **de probabilidades**, de los estimadores de MCO. Sin eso no es posible realizar ningún tipo de prueba de hipótesis referente a los valores verdaderos de estos estimadores. Como se muestra en los capítulos 3 y 7, los

estimadores de MCO son funciones lineales de la variable dependiente $Y$, y la propia $Y$ es una función lineal del término de error estocástico $u$, suponiendo que las variables explicativas son no estocásticas, o fijas, en muestras repetidas. En última instancia, necesitamos la distribución de probabilidades de $u$.

Como se mencionó, el *modelo clásico de regresión lineal normal* (MCRLN) supone que el término de error sigue la distribución normal (con media cero y varianza constante). Aplicamos el **teorema central del límite (TCL)** para justificar la normalidad del término de error y demostramos que los estimadores de MCO están distribuidos normalmente con medias y varianza analizadas en los capítulos 4 y 7. Esto, a su vez, permitió usar los estadísticos $t$ y $F$ para probar hipótesis en muestras pequeñas, o finitas, así como en muestras grandes. Por tanto, la función del supuesto de normalidad es crucial, en especial en muestras pequeñas.

Pero, ¿qué pasa si no es posible sostener el supuesto de normalidad con base en las diferentes pruebas de normalidad? ¿Y después qué? Existen dos opciones. La primera es el **remuestreo**, y la segunda consiste en invocar la **teoría de muestras grandes**, o **asintóticas**.

El análisis del método de remuestreo, que poco a poco se filtra en la econometría aplicada, nos desviaría mucho del tema. La idea básica del remuestreo es *agitar* (o revolver) una muestra dada una y otra vez y luego obtener las distribuciones muestrales de los parámetros de interés (estimadores de MCO, para estos propósitos). El método en la práctica se deja como tema de consulta.[52] A propósito, el término procede de la expresión popular "salir adelante por uno mismo".

El otro método para tratar con los términos de error no normales es aplicar la teoría de muestras grandes, o asintóticas. De hecho, dimos un vistazo a esto en el apéndice 3A.7, en el capítulo 3, cuando demostramos que los estimadores de MCO son **consistentes**. Como analizamos en el **apéndice A**, un estimador es consistente si se aproxima al valor verdadero del estimador conforme la muestra aumenta cada vez más (véase la figura A.11, en el apéndice A).

Pero, ¿de qué sirve eso en las pruebas de hipótesis? ¿Es posible seguir usando las pruebas $t$ y $F$? Se puede demostrar que según los supuestos de Gauss-Markov, los estimadores de MCO tienen **distribución normal asintótica**, con las medias y varianzas que expusimos en los capítulos 4 y 7.[53] Como resultado, las pruebas $t$ y $F$ ejecutadas según el supuesto de normalidad son *aproximadamente* válidas en muestras grandes. La aproximación llega a ser muy buena conforme aumenta el tamaño de la muestra.[54]

## 2. Variables explicativas estocásticas

En el capítulo 3 introdujimos el modelo clásico de regresión lineal (en los parámetros) sgún ciertos supuestos de simplificación. Un supuesto fue que las variables explicativas, o regresoras, eran fijas o no estocásticas, o, si eran estocásticas, eran independientes del término de error. El primer caso se denomina *caso de regresoras fijas*, y el segundo, *de regresoras aleatorias*.

---

[52] Para un análisis informal, véase Christopher Z. Mooney y Robert D. Duval, *Bootstrapping: A Nonparametric Approach to Statistical Inference,* Sage University Press, California, 1993. Para un análisis clásico más formal, véase Russell Davidson y James G. MacKinnon, *Econometric Theory and Methods,* Oxford University Press, Nueva York, 2004, pp. 159-166.

[53] Recuerde los supuestos de Gauss-Markov, a saber: el valor esperado del término de error es cero, el término de error y cada una de las variables explicativas son independientes, la varianza del error es homoscedástica y no hay autocorrelación en el término de error. También se supone que la matriz de varianza-covarianza de las variables explicativas es finita. Asimismo, podemos flexibilizar la condición de independencia entre el término de error y las regresoras, y suponer la condición más débil de que no están correlacionadas.

[54] La prueba de la normalidad asintótica de los estimadores de MCO trasciende el alcance de este libro. Véase James H. Stock y Mark W. Watson, *Introduction to Econometrics,* 2a. ed., Pearson/Addison Wesley, Boston, 2007, pp. 710-711.

En el *caso de regresoras fijas* ya conocemos las propiedades de los estimadores de MCO (véanse los capítulos 5 y 8). En el *caso de regresoras aleatorias*, si mantenemos el supuesto de que el análisis *depende* de los valores dados de las regresoras, las propiedades de los estimadores de MCO que estudiamos en el caso de las regresoras fijas siguen siendo válidas.

Si en el caso de regresoras aleatorias suponemos que dichas regresoras y el término de error están distribuidos de manera independiente, los estimadores de MCO siguen siendo insesgados, pero no son eficientes.[55]

Las cosas se complican si el término de error no está distribuido normalmente, o las regresoras son estocásticas, o ambas cosas. Aquí es difícil generalizar sobre las propiedades de muestras finitas de los estimadores de MCO. Sin embargo, en ciertas condiciones podemos invocar el teorema central del límite para establecer la normalidad asintótica de los estimadores de MCO. Aunque están fuera del ámbito de este libro, hay demostraciones en otros textos.[56]

## 13.13 Advertencia para el profesional

Hemos visto una enorme cantidad de cosas en este capítulo. No hay duda de que la elaboración de modelos es tanto un arte como una ciencia. Un investigador práctico quizá se desconcierte por las sutilezas teóricas y el conjunto de herramientas de diagnóstico. Pero conviene tener en mente la advertencia de Martin Feldstein: "El econometrista aplicado, como el teórico, pronto descubre a partir de la experiencia que un modelo útil no es el 'verdadero' o 'real', sino el escueto, factible e informativo".[57]

Peter Kennedy, de la Universidad Simon Fraser de Canadá, establece los siguientes "diez mandamientos de la econometría aplicada":[58]

1. Utilizarás el sentido común y la teoría económica.
2. Plantearás las preguntas adecuadas (es decir, preferirás la relevancia antes que la elegancia matemática).
3. Conocerás el contexto (no harás análisis estadísticos ignorantes).
4. Inspeccionarás los datos.
5. No idolatrarás la complejidad. Utilizarás el **principio MSE**; es decir, *mantener la simplicidad estocástica*.
6. Verás las consecuencias de tus resultados y serás perseverante con ellos.
7. Estarás consciente de los costos de la minería de datos.
8. Estarás dispuesto a comprometerte (no adorarás las prescripciones de los libros de texto).
9. No confundirás significancia con sustancia (no confundirás la significancia estadística con la significancia práctica).
10. Te confesarás ante la presencia de la sensibilidad (es decir, te anticiparás a las críticas).

Tal vez el lector desee consultar la totalidad del escrito de Kennedy para apreciar la convicción con la que defiende los diez mandamientos anteriores. Algunos pueden parecer sarcásticos, pero hay más de un grano de verdad en cada uno de ellos.

[55] Para los detalles técnicos, véase William H. Greene, *Econometric Analysis,* 6a. ed., Pearson/Prentice-Hall, Nueva Jersey, 2008, pp. 49-50.

[56] Véase Greene, *op. cit.*

[57] Martin S. Feldstein, "Inflation, Tax Rules and Investment: Some Econometric Evidence", *Econometrica,* vol. 30, 1982, p. 829.

[58] Peter Kennedy, *op. cit.,* pp. 17-18.

## Resumen y conclusiones

1. El supuesto del MCRL de que el modelo econométrico del análisis está correctamente especificado tiene dos significados. Primero, que no hay **errores de especificación ecuacionales** y segundo, que no hay **errores de especificación de modelo**. En este capítulo, el enfoque principal fueron los errores de especificación ecuacionales.
2. Los errores de especificación ecuacionales analizados en este capítulo fueron: 1) omisión de una(s) variable(s) importante(s), 2) inclusión de una(s) variable(s) superflua(s), 3) adopción de la forma funcional equivocada, 4) especificación incorrecta del término de error $u_i$ y 5) errores de medición en la variable regresada y en las regresoras.
3. Cuando se omiten variables legítimas del modelo, las consecuencias pueden ser muy graves: los estimadores de MCO de las variables consideradas en el modelo no sólo están sesgados sino que también son inconsistentes. Además, las varianzas y los errores estándar de estos coeficientes están estimados en forma incorrecta, lo que vicia los procedimientos usuales de pruebas de hipótesis.
4. Afortunadamente, las consecuencias de incluir variables irrelevantes en el modelo son menos graves: los estimadores de los coeficientes de las variables relevantes, al igual que los de las variables "irrelevantes", permanecen insesgados y continúan siendo consistentes, y la varianza del error $\sigma^2$ permanece correctamente estimada. El único problema es que las varianzas estimadas tienden a ser más grandes de lo necesario, lo que resta precisión a la estimación de los parámetros. Es decir, los intervalos de confianza tienden a ser más grandes de lo necesario.
5. Para detectar los errores de especificación ecuacional consideramos diversas pruebas, como: 1) examen de residuos, 2) estadístico *d* de Durbin-Watson, 3) Prueba RESET de Ramsey y 4) prueba del multiplicador de Lagrange.
6. Una clase especial de error de especificación son los errores de medición en los valores de la variable regresada y de las regresoras. Si hay errores de medición sólo en la variable regresada, los estimadores de MCO son insesgados y consistentes, pero menos eficientes. Si hay errores de medición en las regresoras, los estimadores de MCO son sesgados e inconsistentes.
7. Aunque se detecten o sospeche de errores de medición, a menudo las medidas correctivas no son fáciles. Las variables instrumentales o representantes son teóricamente atractivas, pero no siempre prácticas. Por tanto, es muy importante en la vida real que el investigador tenga cuidado al establecer las fuentes de su información, al conocer la forma en que se obtuvo, sus definiciones, etc. La información recolectada por agencias oficiales suele presentarse con diversas notas al pie y el investigador debe advertir al lector de su existencia.
8. Los errores de una mala especificación del modelo pueden ser tan graves como los errores de especificación ecuacionales. En particular, distinguimos entre modelos anidados y no anidados. Para decidir el modelo apropiado analizamos la prueba *F* anidada, o incluyente, así como la prueba *J* de Davidson-MacKinnon, y señalamos las limitaciones de cada una.
9. Al elegir un modelo empírico en la práctica, los investigadores utilizan una variedad de criterios, de los cuales analizamos algunos, como los de información de Akaike y el de Schwarz, el criterio $C_p$ de Mallows y el pronóstico $\chi^2$. Estudiamos las ventajas y desventajas de estos criterios y también advertimos al lector que no eran absolutos, sino complementarios de un análisis de especificación cuidadoso.
10. También analizamos estos temas adicionales: 1) valores atípicos, apalancamientos e influencias, 2) mínimos cuadrados recursivos y 3) prueba de la falla de predicción de Chow. Examinamos el papel de cada uno en el trabajo aplicado.
11. Analizamos brevemente dos casos especiales, a saber, la no normalidad del término de error estocástico y las regresoras aleatorias, y la función de la teoría de muestras grandes, o asintóticas, en situaciones en que no se pueden establecer las propiedades de muestras pequeñas, o finitas, de los estimadores de MCO.

12. Concluimos con el estudio de los "diez mandamientos de la econometría aplicada", de Peter Kennedy. La idea es exigir al investigador que trascienda los aspectos puramente técnicos de la econometría.

## EJERCICIOS

### Preguntas

13.1. Consulte la función de demanda de pollos estimada en la ecuación (8.6.23). Considerando los atributos de un buen modelo analizados en la sección 13.1, ¿puede decir que esta función de demanda está especificada "correctamente"?

13.2. Suponga que el verdadero modelo es

$$Y_i = \beta_1 X_i + u_i \tag{1}$$

pero, en lugar de especificar esta regresión a través del origen, especifica el modelo usual con presencia del intercepto:

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i \tag{2}$$

Evalúe las consecuencias de este error de especificación.

13.3. Continúe con el ejercicio 13.2, pero suponga que el modelo (2) es el verdadero. Analice las consecuencias de ajustar el modelo mal especificado (1).

13.4. Suponga que el "verdadero" modelo es

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_t \tag{1}$$

pero añadimos una variable "irrelevante", $X_3$, al modelo (irrelevante en el sentido de que el verdadero coeficiente $\beta_3$ que acompaña a la variable $X_3$ es cero) y estime

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i \tag{2}$$

*a*) ¿$R^2$ y $R^2$ ajustada para el modelo (2) serían más grandes que las del modelo (1)?

*b*) ¿Las estimaciones de $\beta_1$ y $\beta_2$ obtenidas de (2) son insesgadas?

*c*) ¿La inclusión de la variable "irrelevante" $X_3$ afecta las varianzas de $\hat{\beta}_1$ y $\hat{\beta}_2$?

13.5. Considere la siguiente función de producción (Cobb-Douglas) "verdadera":

$$\ln Y_i = \alpha_0 + \alpha_1 \ln L_{1i} + \alpha_2 \ln L_{2i} + \alpha_3 \ln K_i + u_i$$

donde $Y =$ producción
$L_1 =$ trabajo contenido en la producción
$L_2 =$ trabajo no contenido en la producción
$K =$ capital

Pero suponga que la regresión realmente utilizada en la investigación empírica es

$$\ln Y_i = \beta_0 + \beta_1 \ln L_{1i} + \beta_2 \ln K_i + u_i$$

Conforme al supuesto de que tiene información de corte transversal sobre las variables relevantes,

*a*) ¿Será $E(\hat{\beta}_1) = \alpha_1$ y $E(\hat{\beta}_2) = \alpha_3$?

*b*) Si se sabe que $L_2$ es un insumo *irrelevante* en la función de producción, ¿aún es válida la respuesta en *a*)? Muestre las derivaciones necesarias.

13.6. Consulte las ecuaciones (13.3.4) y (13.3.5). Como se ve, $\hat{\alpha}_2$, aunque sesgada, tiene una varianza menor que $\hat{\beta}_2$, que es insesgada. ¿Cómo decidiría respecto de un intercambio de

un sesgo por una varianza pequeña? *Sugerencia:* El ECM (error cuadrático medio) para los dos estimadores se expresa como

$$\text{ECM}(\hat{\alpha}_2) = \left(\sigma^2 \Big/ \sum x_{2i}^2\right) + \beta_3^2 b_{32}^2$$
$$= \text{varianza muestral} + \text{sesgo al cuadrado}$$
$$\text{ECM}(\hat{\beta}_2) = \sigma^2 \Big/ \sum x_2^2 \left(1 - r_{23}^2\right)$$

Respecto de ECM, véase el **apéndice A**.

13.7. Muestre que el $\beta$ estimado de (13.5.1) o (13.5.3) constituye una estimación insesgada del verdadero $\beta$.

13.8. Según la hipótesis de ingreso permanente de Friedman, podemos escribir

$$Y_i^* = \alpha + \beta X_i^* \qquad (1)$$

donde $Y_i^*$ = gasto de consumo "permanente" y $X_i^*$ = ingreso "permanente". En lugar de las variables "permanentes", observamos

$$Y_i = Y_i^* + u_i$$
$$X_i = X_i^* + v_i$$

donde $Y_i$ y $X_i$ son las cantidades observables o mensurables, y donde $u_i$ y $v_i$ son los errores de medición en $Y^*$ y $X^*$, respectivamente.

Con las cantidades observables, escribimos la función de consumo como

$$Y_i = \alpha + \beta(X_i - v_i) + u_i$$
$$= \alpha + \beta X_i + (u_i - \beta v_i) \qquad (2)$$

Si suponemos que 1) $E(u_i) = E(v_i) = 0$, 2) var $(u_i) = \sigma_u^2$ y var $(v_i) = \sigma_v^2$, 3) cov$(Y_i^*, u_i) = 0$, cov$(X_i^*, v_i) = 0$, y 4) cov$(u_i, X_i^*) =$ cov$(v_i, Y_i^*) =$ cov$(u_i, v_i) = 0$, demuestre que, en muestras grandes, el $\beta$ estimado de (2) se expresa como

$$\text{plím}\,(\hat{\beta}) = \frac{\beta}{1 + \left(\sigma_v^2/\sigma_{X^*}^2\right)}$$

*a*) ¿Qué puede decir sobre la naturaleza del sesgo en $\hat{\beta}$?

*b*) Si el tamaño de la muestra aumenta indefinidamente, ¿tenderá el $\beta$ estimado a igualar el $\beta$ verdadero?

13.9. *Modelo de asignación de precios de activos de capital*. El modelo de asignación de precios de activos de capital (CAPM) de la teoría de inversión moderna postula la siguiente relación entre la tasa de rendimiento promedio de un valor (una acción común), medida durante un determinado periodo, y la volatilidad del título, relación denominada *coeficiente Beta* (la volatilidad es una medida del riesgo):

$$\bar{R}_i = \alpha_1 + \alpha_2(\beta_i) + u_i \qquad (1)$$

donde $\bar{R}_i$ = tasa de rendimiento promedio del valor $i$
$\beta_i$ = coeficiente Beta verdadero del valor $i$
$u_i$ = término de perturbación estocástico

La verdadera $\beta_i$ no es directamente observable pero se mide de la siguiente manera:

$$r_{it} = \alpha_1 + \beta^* r_{m_t} + e_t \qquad (2)$$

donde $r_{it} =$ tasa de rendimiento del valor $i$ durante el tiempo $t$

$r_{mt} =$ tasa de rendimiento del mercado durante el tiempo $t$ (esta tasa es la tasa de rendimiento sobre algún índice general del mercado, como el índice S&P de valores industriales)

$e_t =$ término de residuos

y donde $\beta^*$ es una estimación del "verdadero" coeficiente beta. En la práctica, por consiguiente, en lugar de estimar (1), se estima

$$\bar{R}_i = \alpha_1 + \alpha_2(\beta_i^*) + u_i \tag{3}$$

donde las $\beta_i^*$ se obtienen de la regresión (2). Pero, como las $\beta_i^*$ son estimadas, la relación entre la verdadera $\beta$ y $\beta^*$ se escribe así

$$\beta_i^* = \beta_i + v_i \tag{4}$$

donde $v_i$ se denomina *error de medición*.

*a*) ¿Cuál será el efecto de este error de medición sobre la estimación de $\alpha_2$?

*b*) ¿El $\alpha_2$ estimado de (3) proporcionará una estimación insesgada del verdadero $\alpha_2$? De no ser así, ¿es ésta una estimación consistente de $\alpha_2$? De no ser así, ¿qué medidas correctivas sugiere?

13.10. Considere el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \tag{1}$$

Para averiguar si este modelo está mal especificado porque omite la variable $X_3$, decide efectuar la regresión de los residuos obtenidos del modelo (1) sólo sobre la variable $X_3$ (*Nota:* Hay un intercepto en esta regresión.) La prueba del multiplicador de Lagrange (ML), sin embargo, requiere la regresión de los residuos de (1) sobre $X_2$, $X_3$ y una constante. ¿Por qué es probable que su procedimiento sea inapropiado?*

13.11. Considere el modelo

$$Y_i = \beta_1 + \beta_2 X_i^* + u_i$$

En la práctica, medimos $X_i^*$ mediante $X_i$ de manera que

*a*) $X_i = X_i^* + 5$

*b*) $X_i = 3X_i^*$

*c*) $X_i = (X_i^* + \varepsilon_i)$, donde $\varepsilon_i$ es un término puramente aleatorio con las propiedades usuales.

¿Cuál será el efecto de estos errores de medición sobre las estimaciones de los verdaderos $\beta_1$ y $\beta_2$?

13.12. Consulte las ecuaciones de regresión (13.3.1) y (13.3.2). En forma similar a (13.3.3), demuestre que

$$E(\hat{\alpha}_1) = \beta_1 + \beta_3(\bar{X}_3 - b_{32}\bar{X}_2)$$

donde $b_{3\,2}$ es el coeficiente de pendiente en la regresión de la variable omitida $X_3$ sobre la variable incluida $X_2$.

13.13. Evalúe de manera crítica el siguiente punto de vista, de Leamer:†

---

* Véase Maddala, *op. cit.*, p. 477.

† Edward E. Leamer, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons, Nueva York, 1978, p. vi.

> Mi interés en la metaestadística [es decir, en la teoría de la inferencia obtenida realmente de los datos] surge de mis observaciones del trabajo de los economistas. La opinión de que la teoría econométrica es irrelevante la sostiene una enorme y vergonzosa cantidad de profesionales de la economía. Es de esperarse que el enorme abismo entre la teoría y la práctica de la econometría provoque tensiones profesionales. De hecho, un balance ecuánime permea nuestras publicaciones y nuestras reuniones. Nos dividimos cómodamente en un sacerdocio célibe de teóricos estadísticos, por una parte, y una legión de incorregibles pecadores analistas de datos, por otra. Los sacerdotes tienen el poder de elaborar listas de pecados y son reverenciados por los talentos especiales que ostentan. No se espera que los pecadores dejen de pecar, sólo se necesita que confiesen sus errores públicamente.

13.14. Evalúe la siguiente afirmación de Henry Theil:*

> En el actual nivel técnico, el procedimiento más sensible es interpretar los coeficientes de confianza y los límites de significancia de manera liberal, cuando los intervalos de confianza y los estadísticos de prueba se calculan a partir de la regresión final de una estrategia de regresión, en forma convencional. Es decir, un coeficiente a 95% de confianza en realidad puede ser un coeficiente a 80% de confianza, y un nivel de significancia de 1% de en realidad puede ser de 10 por ciento.

13.15. Al comentar la metodología econométrica practicada en la década de 1950 y principios de la siguiente, Blaug expresó:†

> . . . gran parte de ésta [la investigación empírica] se asemeja a jugar tenis con la red abajo; en lugar de refutar las predicciones que pueden probarse, los economistas modernos con demasiada frecuencia se contentan con demostrar que el mundo real se ajusta a sus predicciones, con lo que remplazan así la falsificación [al estilo Popper], la cual es complicada, con la verificación, que es muy sencilla.

¿Está de acuerdo con lo anterior? Quizá desee consultar la obra de Blaug para conocer más sus puntos de vista.

13.16. De acuerdo con Blaug, "No hay lógica para la prueba, pero sí la hay en la refutación".‡ ¿Qué quiso decir?

13.17. Consulte el modelo de St. Louis analizado en el texto. Tenga en cuenta los problemas relacionados con la prueba $F$ anidada y evalúe de manera crítica los resultados presentados en la regresión (13.8.4).

13.18. Suponga que el modelo verdadero es

$$Y_i = \beta_1 + \beta_2 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

pero estima

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$

Si utiliza las observaciones de $Y$ en $X = -3, -2, -1, 0, 1, 2, 3$, y estima el modelo "incorrecto", ¿qué sesgo se obtendrá en estas estimaciones?§

13.19. Para saber si la variable $X_i^2$ pertenece al modelo $Y_i = \beta_1 + \beta_2 X_i + u_i$, la prueba RESET de Ramsey estimaría el modelo lineal para obtener la estimación de los valores $Y_i$ de este modelo [es decir, $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$], después estimaría el modelo $Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 \hat{Y}_i^2 + v_i$ y luego probaría la significancia de $\alpha_3$. Demuestre que si $\hat{\alpha}_3$ resulta estadísticamente significativa en la ecuación anterior (RESET), equivale a estimar el siguiente

* Henry Theil, *Principles of Econometrics*, John Wiley & Sons, Nueva York, 1971, pp. 605-606.

† M. Blaug, *The Methodology of Economics. Or How Economists Explain*, Cambridge University Press, Nueva York, 1980, p. 256.

‡ Ibid., p. 14.

§ Adaptado de G.A.F., *Linear Regression Analysis*, John Wiley & Sons, Nueva York, 1977, p. 176.

modelo de manera directa: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$. (*Sugerencia:* Sustituya por $\hat{Y}_i$ en la regresión RESET.)*

13.20. Fundamente con argumentos si los siguientes enunciados son verdaderos o falsos.†

*a*) Una observación puede ser influyente pero no ser un valor atípico.

*b*) Una observación puede ser un valor atípico, pero no ser influyente.

*c*) Una observación puede ser un valor atípico e influyente.

*d*) Si el modelo $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i \hat{\beta}_3$ resulta estadísticamente significativo, se debe conservar el término lineal $X_i$ aunque $\hat{\beta}_2$ sea estadísticamente insignificante.

*e*) Si estima el modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ o $Y_i = \alpha_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ mediante MCO, la línea de regresión estimada es la misma, donde $(X_{2i} - \bar{X}_2)$ y $x_{3i} = (X_{3i} - \bar{X}_3)$.

## Ejercicios empíricos

13.21. Utilice la información de la demanda de pollos del ejercicio 7.19. Suponga que la verdadera función de demanda es

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_6 \ln X_{6t} + u_t \tag{1}$$

pero considera la siguiente función de demanda:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + v_t \tag{2}$$

donde $Y =$ consumo de pollos *per cápita* (en libras)

$X_2 =$ ingreso real disponible *per cápita*

$X_3 =$ precio real de los pollos al menudeo

$X_6 =$ precio real compuesto de los sustitutos del pollo

*a*) Realice las pruebas RESET y ML de errores de especificación, suponiendo que la función de demanda (1) dada es la verdadera.

*b*) Suponga que $\hat{\beta}_6$ en (1) resulta estadísticamente no significativa. ¿Indica esto que no hay error de especificación si se ajusta (2) a la información?

*c*) Si $\hat{\beta}_6$ resulta no significativa, ¿indica eso que no se debe introducir el precio de uno o más productos sustitutos como argumento en la función de demanda?

13.22. Continúe con el ejercicio 13.21. Estrictamente por razones pedagógicas, suponga que el modelo (2) es la verdadera función de demanda.

*a*) Si ahora estimamos el modelo (1), ¿qué tipo de error de especificación se comete en esta instancia?

*b*) ¿Cuáles son las consecuencias teóricas de este error de especificación? Ilustre con la información disponible.

13.23. El verdadero modelo es

$$Y_i^* = \beta_1 + \beta_2 X_i^* + u_i \tag{1}$$

pero, debido a errores de medición, estima

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i \tag{2}$$

donde $Y_i = Y_i^* + \varepsilon_i$ y $X_i = X_i^* + w_i$, donde $\varepsilon_i$ y $w_i$ son errores de medición.

---

* Adaptado de Kerry Peterson, *op. cit.*, pp. 184-185.

† Adaptado de Norman R. Draper y Harry Smith, *op. cit.*, pp. 606-607.

Con la información de la tabla 13.2, evalúe las consecuencias de estimar (2) en lugar del verdadero modelo (1).

13.24. *Experimento Monte Carlo:** Diez personas percibieron el siguiente ingreso permanente semanal: $200, 220, 240, 260, 280, 300, 320, 340, 380 y 400. El consumo permanente ($Y_i^*$) estuvo relacionado con el ingreso permanente, $X_i^*$, así

$$Y_i^* = 0.8X_i^* \tag{1}$$

Cada individuo tuvo un ingreso transitorio igual a 100 veces un número aleatorio $u_i$, obtenido de una población normal con media = 0 y $\sigma^2 = 1$ (es decir, una variable normal estandarizada). Suponga que no hay un componente transitorio en el consumo. Así, el consumo medido y el consumo permanente son los mismos.

*a*) Seleccione 10 números aleatorios de una población normal con media cero y varianza unitaria, y obtenga 10 números para el ingreso medido $X_i (= X_i^* + 100u_i)$.

*b*) Efectúe la regresión del consumo permanente (= medido) sobre el ingreso medido con la información obtenida en *a*) y compare los resultados con los que aparecen en (1). *A priori*, el intercepto debe ser cero (¿por qué?) ¿Es ése el caso? ¿Por qué?

*c*) Repita *a*) 100 veces, obtenga 100 regresiones como las de *b*) y compare los resultados con la verdadera regresión (1). ¿Qué conclusiones generales obtiene?

13.25. Consulte el ejercicio 8.26. Con las definiciones de las variables dadas ahí, considere los dos modelos siguientes para explicar $Y$:

$$\text{Modelo A:} \quad Y_t = \alpha_1 + \alpha_2 X_{3t} + \alpha_3 X_{4t} + \alpha_4 X_{6t} + u_t$$

$$\text{Modelo B:} \quad Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{5t} + \beta_4 X_{6t} + u_t$$

Con la prueba $F$ anidada, ¿cómo elegiría alguno de los dos modelos?

13.26. Continúe con el ejercicio 13.25. Con la prueba $J$, ¿cómo escogería alguno de los dos modelos?

13.27. Consulte el ejercicio 7.19, que analiza la demanda de pollo en Estados Unidos. Ahí se dan cinco modelos.

*a*) ¿Cuál es la diferencia entre el modelo 1 y el 2? Si el modelo 2 es correcto y estima el modelo 1, ¿qué tipo de error comete? ¿Qué pruebas aplicaría: para el error de especificación de ecuación, o para el error de selección de modelo? Muestre los cálculos necesarios.

*b*) Entre los modelos 1 y 5, ¿cuál elegiría? ¿Qué prueba(s) aplicaría(n) y por qué?

13.28. Consulte la tabla 8.11, que proporciona información sobre los ahorros personales ($Y$) y el ingreso personal disponible ($X$) de 1970 a 2005. Ahora considere los siguientes modelos:

$$\text{Modelo A:} \quad Y_t = \alpha_1 + \alpha_2 X_t + \alpha_3 X_{t-1} + u_t$$

$$\text{Modelo B:} \quad Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

¿Cómo eligiría alguno de los dos modelos? Establezca con claridad el (los) procedimiento(s) de prueba que utilice y muestre todos los cálculos. Suponga que se cuestiona que la variable tasa de interés pertenece a la función ahorro: ¿cómo probaría eso? Recopile los datos de la tasa de interés para bonos del tesoro a tres meses como un representante de los intereses y justifique su respuesta.

* Adaptado de Christopher Dougherty, *Introduction to Econometrics,* Oxford University Press, Nueva York, 1992, pp. 253-256.

13.29. Utilice los datos del ejercicio 13.28. Para familiarizarse con los mínimos cuadrados recursivos, calcule las funciones ahorro de 1970-1981, 1970-1985, 1970-1990 y 1970-1995. Comente la estabilidad de los coeficientes estimados en las funciones ahorro.

13.30. Continúe con el ejercicio 13.29, pero ahora use los datos actualizados de la tabla 8.10.

*a*) Suponga que estima la función de ahorro de 1970-1981. Con los parámetros así estimados y los datos del ingreso personal disponible de 1982-2000, estime el ahorro pronosticado para el segundo periodo y use la prueba de falla de predicción de Chow para averiguar si se rechaza la hipótesis de que la función de ahorro entre los dos periodos no ha cambiado.

*b*) Ahora estime la función de ahorro de los datos de 2000-2005. Compare los resultados con la función correspondiente al periodo 1982-2000 mediante el mismo método que en el inciso anterior (la prueba de falla de predicción de Chow). ¿Hay algún cambio significativo en la función de ahorro entre los dos periodos?

13.31. *Omisión de una variable en el modelo de regresión con K variables*. Consulte la ecuación (13.3.3), que muestra el sesgo por omitir la variable $X_3$ del modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$. Esto se generaliza de la siguiente forma: en el modelo con $k$ variables, $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, suponga que omitimos la variable $X_k$. Entonces, es posible demostrar que el sesgo de la variable omitida que corresponde al coeficiente de la pendiente para la variable incluida $X_j$ es:

$$E(\hat{\beta}_j) = \beta_j + \beta_k b_{kj} \qquad j = 2, 3, \ldots, (k-1)$$

donde $b_{kj}$ es el coeficiente de la pendiente (parcial) de $X_j$ en la regresión auxiliar de la variable excluida $X_k$ sobre todas las variables explicativas incluidas en el modelo.*

Consulte el ejercicio 13.21. Obtenga el sesgo de los coeficientes en la ecuación (1) si excluimos la variable ln $X_6$ del modelo. ¿Esta exclusión es grave? Muestre los cálculos necesarios.

# Apéndice 13A

## 13A.1 Prueba de que $E(b_{12}) = \beta_2 + \beta_3 b_{32}$ [ecuación (13.3.3)]

En la forma de desviación, el modelo de regresión de población con tres variables se expresa

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + (u_i - \bar{u}) \tag{1}$$

Si primero se multiplica por $x_2$ y luego por $x_3$, las ecuaciones normales usuales son:

$$\sum y_i x_{2i} = \beta_2 \sum x_{2i}^2 + \beta_3 \sum x_{2i} x_{3i} + \sum x_{2i}(u_i - \bar{u}) \tag{2}$$

$$\sum y_i x_{3i} = \beta_2 \sum x_{2i} x_{3i} + \beta_3 \sum x_{3i}^2 + \sum x_{3i}(u_i - \bar{u}) \tag{3}$$

Al dividir (2) entre $\sum x_{2i}^2$ en ambos lados, obtenemos

$$\frac{\sum y_i x_{2i}}{\sum x_{2i}^2} = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{4}$$

* Lo anterior se generaliza al caso en el que más de una variable *X* relevante esté excluida del modelo. Sobre este tema, véase Chandan Mukherjee *et al., op. cit.*, p. 215.

Ahora, si recuerda que

$$b_{12} = \frac{\sum y_i x_{2i}}{\sum x_{2i}^2}$$

$$b_{32} = \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2}$$

la ecuación (4) se expresa como

$$b_{12} = \beta_2 + \beta_3 b_{32} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{5}$$

Si toma el valor esperado de (5) en ambos lados, obtenemos finalmente

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \tag{6}$$

donde se aprovecha que: *a*) para una muestra dada, $b_{32}$ es una cantidad fija conocida, *b*) $\beta_2$ y $\beta_3$ son constantes y c) $u_i$ no está correlacionada con $X_{2i}$ (ni tampoco con $X_{3i}$).

## 13A.2 Consecuencias de la inclusión de una variable irrelevante: propiedad de insesgamiento

Para el verdadero modelo (13.3.6), tenemos

$$\hat{\beta}_2 = \frac{\sum y x_2}{\sum x_2^2} \tag{1}$$

y sabemos que es insesgado.

Para el modelo (13.3.7) obtenemos

$$\hat{\alpha}_2 = \frac{\left(\sum y x_2\right)\left(\sum x_3^2\right) - \left(\sum y x_3\right)\left(\sum x_2 x_3\right)}{\sum x_2^2 \sum x_3^2 - \left(\sum x_2 x_3\right)^2} \tag{2}$$

Ahora el verdadero modelo en forma de desviación es

$$y_i = \beta_2 x_2 + (u_i - \bar{u}) \tag{3}$$

Al sustituir para $y_i$ de (3) en (2) y simplificar, obtenemos

$$E(\hat{\alpha}_2) = \beta_2 \frac{\sum x_2^2 \sum x_3^2 - \left(\sum x_2 x_3\right)^2}{\sum x_2^2 \sum x_3^2 - \left(\sum x_2 x_3\right)^2} \tag{4}$$

$$= \beta_2$$

es decir, $\hat{\alpha}_2$ permanece insesgado.

También obtenemos

$$\hat{\alpha}_3 = \frac{\left(\sum y x_3\right)\left(\sum x_2^2\right) - \left(\sum y x_2\right)\left(\sum x_2 x_3\right)}{\sum x_2^2 \sum x_3^2 - \left(\sum x_2 x_3\right)^2} \tag{5}$$

Al sustituir para $y_i$ de (3) en (5) y simplificar, obtenemos

$$\begin{aligned} E(\hat{\alpha}_3) &= \beta_2 \frac{\left[\left(\sum x_2 x_3\right)\left(\sum x_2^2\right) - \left(\sum x_2 x_3\right)\left(\sum x_2^2\right)\right]}{\sum x_2^2 \sum x_3^2 - \left(\sum x_2 x_3\right)^2} \\ &= 0 \end{aligned} \tag{6}$$

el cual es su valor en el verdadero modelo, pues $X_3$ está ausente de dicho modelo.

## 13A.3 Prueba de la ecuación (13.5.10)

Tenemos

$$Y = \alpha + \beta X_i^* + u_i \tag{1}$$

$$X_i = X_i^* + w_i \tag{2}$$

Por consiguiente, en la forma de desviación, obtenemos

$$y_i = \beta x_i^* + (u_i - \bar{u}) \tag{3}$$

$$x_i = x_i^* + (w_i - \bar{w}) \tag{4}$$

Ahora, cuando utilizamos

$$Y_i = \alpha + \beta X_i + u_i \tag{5}$$

obtenemos

$$\begin{aligned} \hat{\beta} &= \frac{\sum yx}{\sum x^2} \\ &= \frac{\sum[\beta x^* + (u - \bar{u})][x^* + (w - \bar{w})]}{\sum[x^* + (w - \bar{w})]^2} \qquad \text{con (3) y (4)} \\ &= \frac{\beta \sum x^{*2} + \beta \sum x^*(w - \bar{w}) + \sum x^*(u - \bar{u}) + \sum(u - \bar{u})(w - \bar{w})}{\sum x^{*2} + 2 \sum x^*(w - \bar{w}) + \sum(w - \bar{w})^2} \end{aligned}$$

Como no podemos tomar la esperanza de esta expresión porque la esperanza de la razón de dos variables no es igual a la razón de sus esperanzas (*nota:* el operador de esperanzas $E$ es un operador lineal), primero dividimos cada término del numerador y del denominador entre $n$ y obtenemos la probabilidad del límite, plím (véanse los detalles de plím en el **apéndice A**), de

$$\hat{\beta} = \frac{(1/n)\left[\beta \sum x^{*2} + \beta \sum x^*(w - \bar{w}) + \sum x^*(u - \bar{u}) + \sum(u - \bar{u})(w - \bar{w})\right]}{(1/n)\left[\sum x^{*2} + 2 \sum x^*(w - \bar{w}) + \sum(w - \bar{w})^2\right]}$$

Ahora, la probabilidad del límite de la razón de dos variables es la razón de sus probabilidades del límite. Al aplicar esta regla y tomar el plím de cada término, obtenemos

$$\text{plím}\, \hat{\beta} = \frac{\beta \sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_w^2}$$

donde $\sigma_{X^*}^2$ y $\sigma_w^2$ son las varianzas de $X^*$ y $w$ a medida que el tamaño de la muestra aumenta indefinidamente y donde aprovechamos que, a medida que el tamaño de la muestra aumenta indefinidamente, no hay correlación entre los errores $u$ y $w$ ni entre ellos y la verdadera $X^*$. De la expresión anterior, finalmente obtenemos

$$\text{plím}\, \hat{\beta} = \beta \left[\frac{1}{1 + \left(\sigma_w^2 / \sigma_{X^*}^2\right)}\right]$$

que es el resultado requerido.

## 13A.4 Prueba de la ecuación (13.6.2)

Como no hay término de intercepto en el modelo, la estimación para $\alpha$, de acuerdo con la fórmula para la regresión a través del origen, es como sigue:

$$\hat{\alpha} = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{1}$$

Al sustituir por $Y$ del verdadero modelo (13.2.8), obtenemos

$$\hat{\alpha} = \frac{\sum X_i(\beta X_i u_i)}{\sum X_i^2} = \beta \frac{\sum X_i^2 u_i}{\sum X_i^2} \tag{2}$$

La teoría estadística muestra que si ln $u_i \sim N(0, \sigma^2)$, entonces

$$u_i = \text{log normal}\left[e^{\sigma^2/2},\ e^{\sigma^2}\left(e^{\sigma^2-1}\right)\right] \tag{3}$$

Por tanto,

$$\begin{aligned} E(\hat{\alpha}) &= \beta E\left(\frac{\sum X_i^2 u_i}{\sum X_i^2}\right) \\ &= \beta\left(E\frac{\left(X_1^2 u_1 + X_2^2 u_2 + \cdots + X_n^2 u_n\right)}{\sum X_i^2}\right) \\ &= \beta e^{\sigma^2/2}\left(\frac{\sum X_i^2}{\sum X_i^2}\right) = \beta e^{\sigma^2/2} \end{aligned}$$

donde se aprovecha que las $X$ son no estadísticas y cada $u_i$ tiene un valor esperado de $e^{\sigma^2/2}$.

Como $E(\hat{\alpha}) \neq \beta$, $\hat{\alpha}$ es un estimador sesgado de $\beta$.

Parte 3

# Temas de econometría

En la **parte 1** introdujimos el modelo clásico de regresión lineal con todos sus supuestos. En la **parte 2** examinamos en detalle las consecuencias de no satisfacer uno o más supuestos y lo que puede hacerse al respecto. En la **parte 3** estudiaremos algunas técnicas econométricas seleccionadas pero muy comunes. En particular, analizaremos estos temas: 1) modelos de regresión no lineales en los parámetros, 2) modelos de regresión con respuesta cualitativa, 3) modelos de regresión con datos de paneles y 4) modelos econométricos dinámicos.

En el capítulo 14 consideramos modelos *intrínsecamente* no lineales en los parámetros. Con la actual disponibilidad de paquetes de software, ya no es muy difícil estimar tales modelos. Aunque el fundamento matemático puede ser complejo para ciertos lectores, las ideas básicas de los modelos de regresión no lineales en los parámetros se comprenden por intuición. Con ejemplos adecuados, este capítulo muestra cómo estimar e interpretar estos modelos.

En el capítulo 15 analizaremos los modelos de regresión en los que la variable dependiente es cualitativa por naturaleza. Este capítulo, por tanto, complementa al 9, en el cual estudiamos los modelos en que la naturaleza de las variables explicativas era cualitativa. La idea central de este capítulo es el desarrollo de modelos con la variable regresada del tipo sí o no. Como los mínimos cuadrados ordinarios (MCO) plantean diversos problemas para estimar tales modelos, se han elaborado diversas opciones. En este capítulo examinamos dos de ellas, a saber: el **modelo logit** y el **modelo probit**. En este capítulo también analizamos diversas variantes de los modelos con respuesta cualitativa, como el **modelo tobit** y el **modelo de regresión de Poisson**. Además, estudiaremos brevemente varias extensiones de los modelos con respuesta cualitativa, como los modelos **probit ordenado, logit ordenado** y **logit multinomial**.

En el capítulo 16 veremos los **modelos de regresión con datos de paneles**. Tales modelos combinan las series de tiempo y las observaciones transversales. Aunque al combinar las mencionadas observaciones se incrementa el tamaño de la muestra, la estimación de los modelos de regresión con datos de paneles plantea diversos retos. En este capítulo estudiaremos sólo lo esencial de estos modelos y guiaremos al lector hacia los recursos adecuados para un estudio posterior.

En el capítulo 17 consideraremos los modelos de regresión con valores de las variables explicativas para el periodo actual, lo mismo que para periodos pasados o rezagados, además de modelos que incluyen uno o varios valores rezagados de la variable dependiente como una de las

variables explicativas. Estos modelos se denominan, respectivamente, **de rezago distribuido** y **autorregresivos**. Aunque tales modelos son muy útiles en la econometría empírica, su aplicación conlleva algunos problemas especiales de estimación, pues violan uno o más supuestos del modelo clásico de regresión lineal. Consideraremos estos problemas especiales en el contexto de los modelos de Koyck, de expectativas adaptables (EA) y de ajuste parcial. También abordaremos las críticas contra el modelo EA por parte de los defensores de la llamada escuela de expectativas racionales (ER).

# Capítulo 14

# Modelos de regresión no lineales

Este libro se centra, sobre todo, en los modelos de regresión lineal; es decir, modelos lineales en los parámetros o que pueden transformarse para que lo sean. Sin embargo, en ciertas ocasiones, por razones teóricas o empíricas, debemos considerar modelos que no son lineales en los parámetros.[1] En este capítulo analizaremos tales modelos y estudiaremos sus características especiales.

## 14.1 Modelos de regresión intrínsecamente lineales e intrínsecamente no lineales

Cuando comenzamos el análisis de los modelos de regresión lineal en el capítulo 2, establecimos que la principal preocupación en este libro son sobre todo los modelos lineales en los parámetros pero que pueden ser o no lineales en las variables. Si consulta la tabla 2.3, verá que un modelo lineal en los parámetros y en las variables es un modelo de regresión lineal, de igual forma que lo es un modelo lineal en los parámetros pero no lineal en las variables. Por otro lado, si un modelo es no lineal en los parámetros, se trata de un modelo de regresión no lineal (en los parámetros), sin importar que las variables de tal modelo sean lineales o no.

Sin embargo, debe tener cuidado, pues algunos modelos pueden parecer no lineales en los parámetros pero ser **inherente** o **intrínsecamente** lineales, debido a que con una transformación adecuada pueden convertirse en modelos de regresión lineales en los parámetros. Pero si dichos modelos no pueden linealizarse en los parámetros, se les conoce como **modelos de regresión intrínsecamente no lineales**. *A partir de aquí, cuando hablemos de modelos de regresión no lineales, se trata de modelos intrínsecamente no lineales.* Por brevedad los denominaremos **MRNL**.

Para que resulte clara la distinción entre ambos modelos, revisemos los ejercicios 2.6 y 2.7. En el primero, los modelos **a**, **b**, **c** y **e** son modelos de regresión lineales porque son lineales en todos sus parámetros. El modelo **d** es mixto, pues $\beta_2$ es lineal, pero no $\ln \beta_1$. No obstante, si $\alpha = \ln \beta_1$, entonces este modelo es lineal en $\alpha$ y $\beta_2$.

En el ejercicio 2.7, los modelos **d** y **e** son intrínsecamente no lineales porque no hay una forma simple para linealizarlos. El modelo **c** es obviamente una regresión lineal. ¿Y los mode-

[1] Observamos en el capítulo 4 que, según el supuesto del término de error normalmente distribuido, los estimadores de MCO no sólo son MELI, sino también MEI (mejor estimador insesgado) en toda la clase de estimadores, lineales o no lineales. Pero si eliminamos el supuesto de normalidad, como señalan Davidson y MacKinnon, se pueden obtener estimadores no lineales y/o sesgados tal vez más adecuados que los estimadores de MCO. Véase Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, p. 161.

los **a** y **b**? Al tomar los logaritmos en ambos miembros de **a** obtenemos ln $Y_i = \beta_1 + \beta_2 X_i + u_i$, que es lineal en los parámetros. Por consiguiente, el modelo **a** es *intrínsecamente* un modelo de regresión lineal. El modelo **b** es un ejemplo de una **función de distribución (de probabilidad) logística**, que estudiaremos en el capítulo 15. A primera vista, es un modelo de regresión no lineal, a saber,

$$\ln\left(\frac{1 - Y_i}{Y_i}\right) = \beta_1 + \beta_2 X_i + u_i \tag{14.1.1}$$

Por consiguiente, el modelo **b** es intrínsecamente lineal: veremos la utilidad de los modelos como el (14.1.1) en el siguiente capítulo.

Considere ahora la famosa **función de producción Cobb-Douglas (C-D)**. Sea $Y$ = producción, $X_2$ = insumo trabajo y $X_3$ = insumo capital; expresaremos esta función de tres maneras:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{14.1.2}$$

o

$$\ln Y_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \tag{14.1.2a}$$

donde $\alpha = \ln \beta_1$. En consecuencia, con este formato, la función C-D es intrínsecamente lineal.

Ahora considere esta versión de la función C-D:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i \tag{14.1.3}$$

o

$$\ln Y_i = \alpha + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ln u_i \tag{14.1.3a}$$

donde $\alpha = \ln \beta_1$. Este modelo también es lineal en sus parámetros.

Pero ahora considere la siguiente versión de la función C-D:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} + u_i \tag{14.1.4}$$

Como acabamos de notar, las versiones (14.1.2*a*) y (14.1.3*a*) de la función C-D son modelos de regresión intrínsecamente lineales en los parámetros, pero no hay forma de transformar (14.1.4) de manera que el modelo transformado sea lineal en los parámetros.[2] Por tanto, (14.1.4) es un modelo de regresión intrínsecamente no lineal.

Otra función muy conocida, pero intrínsecamente no lineal, es la función de producción con **elasticidad constante de sustitución (ECS)**, que tiene a la función de producción Cobb-Douglas como caso especial. La producción ECS adopta la siguiente forma:

$$Y_i = A\left[\delta K_i^{-\beta} + (1 - \delta)L_i^{-\beta}\right]^{-1/\beta} \tag{14.1.5}$$

donde $Y$ = producción, $K$ = insumo capital, $L$ = insumo trabajo, $A$ = parámetro de escala, $\delta$ = parámetro de distribución ($0 < \delta < 1$) y $\beta$ = parámetro de sustitución ($\beta \geq -1$).[3] Sin importar la forma en que introduzca el término de error estocástico $u_i$ en esta función producción, no hay manera de cambiarla a un modelo de regresión lineal (en los parámetros). Es intrínsecamente un modelo de regresión no lineal.

[2] Si intenta transformar mediante logaritmos el modelo, no será correcto, pues ln $(A + B) \neq \ln A + \ln B$.

[3] Para las propiedades de la función producción ECS, véase Michael D. Intriligator, Ronald Bodkin y Cheng Hsiao, *Econometric Models, Techniques, and Applications*, 2a. ed., Prentice Hall, 1996, pp. 294-295.

## 14.2 Estimación de modelos de regresión lineales y no lineales

Para apreciar la diferencia al estimar modelos de regresión lineales y no lineales, considere los dos modelos siguientes:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{14.2.1}$$

$$Y_i = \beta_1 e^{\beta_2 X_i} + u_i \tag{14.2.2}$$

A estas alturas ya sabe que (14.2.1) es un modelo de regresión lineal, en tanto que (14.2.2) es no lineal. La regresión (14.2.2) se conoce como **modelo de regresión exponencial**, y a menudo se utiliza para medir el crecimiento de una variable, como la población, el PIB o la oferta de dinero.

Suponga que pretendemos estimar los parámetros de los dos modelos mediante MCO. En ese método reducimos la suma de cuadrados residual (SCR), que para el modelo (14.2.1) es:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{14.2.3}$$

donde, como siempre, $\hat{\beta}_1$ y $\hat{\beta}_2$ son los estimadores de MCO de las verdaderas $\beta$. Al diferenciar la expresión anterior respecto de las dos incógnitas obtenemos las **ecuaciones normales** mostradas en (3.1.4) y (3.1.5). Al resolver ambas de manera simultánea obtenemos los estimadores de MCO dados en las ecuaciones (3.1.6) y (3.1.7). Observe con cuidado que en dichas ecuaciones las incógnitas (las $\beta$) están en el miembro izquierdo y los datos conocidos ($X$ y $Y$) en el derecho. Como resultado, obtenemos soluciones explícitas para ambas incógnitas en términos de los datos dados.

Ahora vea lo que pasa si tratamos de reducir la SCR de (14.2.2). Como vimos en el apéndice 14A, sección 14A.1, las ecuaciones normales correspondientes a (3.1.4) y (3.1.5) son las siguientes:

$$\sum Y_i e^{\hat{\beta}_2 X_i} = \beta_1 e^{2\hat{\beta}_2 X_i} \tag{14.2.4}$$

$$\sum Y_i X_i e^{\hat{\beta}_2 X_i} = \hat{\beta}_1 \sum X_i e^{2\hat{\beta}_2 X_i} \tag{14.2.5}$$

A diferencia de las ecuaciones normales para el modelo de regresión lineal, las ecuaciones normales para la regresión no lineal tienen incógnitas (las $\hat{\beta}$) en ambos miembros de las ecuaciones. Como consecuencia, *no podemos obtener soluciones explícitas* de las incógnitas en términos de las cantidades conocidas. Para decirlo de otra forma, las incógnitas están expresadas en términos de ellas mismas y de los datos. Por tanto, aunque apliquemos el método de los mínimos cuadrados para estimar los parámetros de los modelos de regresión no lineal, no obtendremos soluciones explícitas de las incógnitas. A propósito, los MCO aplicados al modelo de regresión no lineal se conocen como **mínimos cuadrados no lineales (MCNL)**. Por tanto, ¿cuál es la solución? Lo veremos en seguida.

## 14.3 Estimación de modelos de regresión no lineales: método de ensayo y error

Para establecer el contexto consideremos un ejemplo concreto. Los datos de la tabla 14.1 se refieren a las comisiones administrativas que un fondo mutualista líder en Estados Unidos paga a sus consultores de inversión por el manejo de sus bienes. Las comisiones dependen del valor neto de los bienes del fondo. Como se ve, mientras más alto sea el valor neto de los bienes, menores serán las comisiones por consultoría, lo cual aparece claramente en la figura 14.1.

**TABLA 14.1**
**Comisiones por consultoría y tamaño de los activos**

| | Comisión, % | Activos* |
|---|---|---|
| 1 | 0.520 | 0.5 |
| 2 | 0.508 | 5.0 |
| 3 | 0.484 | 10 |
| 4 | 0.46 | 15 |
| 5 | 0.4398 | 20 |
| 6 | 0.4238 | 25 |
| 7 | 0.4115 | 30 |
| 8 | 0.402 | 35 |
| 9 | 0.3944 | 40 |
| 10 | 0.388 | 45 |
| 11 | 0.3825 | 55 |
| 12 | 0.3738 | 60 |

* La columna Activos representa el valor neto del activo en miles de millones de dólares.

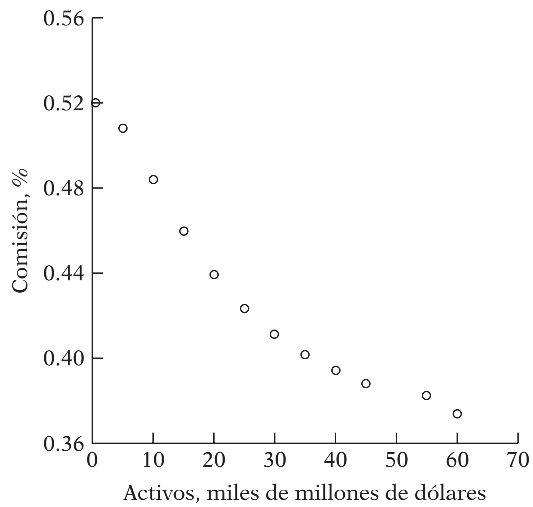



**FIGURA 14.1**
Relación entre la comisión por consultoría y los activos del fondo.

Para ver cómo se ajusta el modelo de regresión exponencial (14.2.2) a los datos de la tabla 14.1, procedemos mediante ensayo y error. Suponga que al principio $\beta_1 = 0.45$ y $\beta_2 = 0.01$. Éstas son sólo conjeturas, a veces fundamentadas en la experiencia previa, el trabajo empírico anterior u obtenidas sólo mediante el ajuste de un modelo de regresión lineal, aunque quizá no resulte el apropiado. En esta etapa no hay que preocuparse por el origen de tales valores.

Como conocemos los valores de $\beta_1$ y $\beta_2$, escribimos la ecuación (14.2.2) como:

$$u_i = Y_i - \beta_1 e^{\beta_2 X_i} = Y_i - 0.45e^{0.01X_i} \quad \textbf{(14.3.1)}$$

Por consiguiente,

$$\sum u_i^2 = \sum (Y_i - 0.45e^{0.01X_i})^2 \quad \textbf{(14.3.2)}$$

Como se conocen $Y$, $X$, $\beta_1$ y $\beta_2$, determinamos con facilidad la *suma de cuadrados de los errores* en (14.3.2).[4] Recuerde que, en MCO, el objetivo es calcular los valores de los parámetros desconocidos que reduzcan lo más posible la suma de cuadrados de los errores. Esto sucede si los

[4] Observe que se conoce a $\sum u_i^2$ como la suma de cuadrados de los errores y no como la usual suma de cuadrados residual, pues se supone que se desconocen los valores de los parámetros.

valores estimados de $Y$ del modelo están tan cerca como sea posible de los valores $Y$ reales. Con los valores dados, obtenemos $\sum u_i^2 = 0.3044$. Pero, ¿cómo sabemos si ésta es la menor suma de cuadrados de los errores que podemos obtener? ¿Qué sucede si elegimos otros valores para $\beta_1$ y $\beta_2$, por ejemplo, 0.50 y $-0.01$, respectivamente? Al repetir el procedimiento, obtenemos ahora $\sum u_i^2 = 0.0073$. Obvio, esta suma de cuadrados de los errores es mucho más pequeña que la anterior, a saber: 0.3044. Pero, ¿cómo sabemos que obtuvimos la suma de cuadrados de los errores más pequeña, y que si eligiéramos otro conjunto de valores para las $\beta$ no obtendríamos otra suma de cuadrados de los errores?

Como se ve, dicho proceso de ensayo y error, o **iterativo**, es sencillo. Y si se dispone de tiempo y paciencia infinitos, el proceso de ensayo y error *quizá* produzca a la larga valores de $\beta_1$ y $\beta_2$ que tal vez garanticen la suma de cuadrados de los errores más pequeña. Pero surge la siguiente pregunta: ¿cómo llegamos de ($\beta_1 = 0.45$; $\beta_2 = 0.01$) a ($\beta_1 = 0.50$; $\beta_2 = -0.1$)? Es evidente que necesitamos una especie de *algoritmo* que indique la forma de ir de un conjunto de valores de las incógnitas a otro conjunto, antes de finalizar. Por fortuna, dichos algoritmos están disponibles, y los analizaremos en la siguiente sección.

## 14.4 Métodos para estimar modelos de regresión no lineales

Existen varios métodos, o algoritmos, para los MRNL: 1) búsqueda directa o método de ensayo y error, 2) optimización directa y 3) linealización iterativa.[5]

### Búsqueda directa o método de ensayo y error, o de libre derivación

En la sección anterior vimos cómo funciona este método. A pesar de ser atractivo, porque se trabaja con la intuición y no requiere cálculo, como otros métodos, no es muy común. *En primer lugar*, si un MRNL contiene varios parámetros, el cálculo de este método se vuelve muy engorroso y pesado. Por ejemplo, si un MRNL implica 5 parámetros y 25 valores alternos para cada parámetro, habrá que calcular la suma de cuadrados de los errores $(25)^5 = 9\ 765\ 625$ veces. *En segundo lugar*, no se garantiza que con el conjunto final de los valores de los parámetros seleccionado necesariamente se obtendrá la suma de cuadrados de los errores mínima absoluta. En el lenguaje del cálculo, obtendría un mínimo local y no un mínimo absoluto. De hecho, ningún método garantiza un mínimo global.

### Optimización directa

En la optimización directa se diferencia la suma de cuadrados de los errores respecto de cada coeficiente o parámetro desconocido, se iguala la ecuación resultante a cero y se resuelven las ecuaciones normales obtenidas de manera simultánea. Ya vimos esto en (14.2.4) y (14.2.5), pero, como se observa a partir de estas ecuaciones, no se pueden resolver explícita o *analíticamente*. Por consiguiente, se requiere alguna rutina iterativa; una de ellas se conoce como **método de pasos descendente**. No veremos los detalles técnicos de este método, pues son un cuanto complejos, pero el lector puede encontrarlos en las referencias. Al igual que el método de ensayo y error, el método de pasos descendente también implica la selección de valores iniciales de prueba para los parámetros desconocidos, pero luego procede de forma más sistemática que el de error y acierto, o el de ensayo y error. Una desventaja de este método es que puede converger a los valores finales de los parámetros de una forma demasiado lenta.

[5] El siguiente análisis se basa en gran medida en las siguientes fuentes: Robert S. Pindyck y Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 4a. ed., McGraw-Hill, 1998, cap. 10; Norman R. Draper y Harry Smith, *Applied Regression Analysis*, 3a. ed., John Wiley & Sons, 1998, cap. 24; Arthur S. Goldberger, *A Course in Econometrics*, Harvard University Press, 1991, cap. 29; Russell Davidson y James MacKinnon, *op. cit.*, pp. 201-207; John Fox, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage, 1997, pp. 393-400, y Ronald Gallant, *Nonlinear Statistical Models*, John Wiley and Sons, 1987.

### Método de linealización iterativa

En este método se linealiza la ecuación no lineal alrededor de algunos valores iniciales de los parámetros. Luego, la ecuación linealizada se calcula mediante MCO y se ajustan los valores elegidos al principio. Con estos valores ajustados se vuelve a *linealizar* el modelo y de nuevo se calcula mediante MCO y se reajustan los valores estimados. Este proceso continúa hasta que no haya un cambio sustancial en los valores estimados respecto del último par de iteraciones. La técnica principal para linealizar una ecuación no lineal es la **expansión de series de Taylor**, tomada del cálculo. En el apéndice 14A, sección 14A.2, se ofrecen detalles muy generales de este método. La estimación del MRNL mediante la expansión de series de Taylor se sistematiza con dos algoritmos conocidos como **método iterativo Gauss-Newton** y **método iterativo Newton-Raphson**. En vista de que uno o ambos métodos ya están incorporados en diversos programas de computadora, y como un estudio de sus aspectos técnicos escapa al objetivo de esta obra, no necesitamos extendernos aquí en su análisis.[6] En la siguiente sección veremos algunos ejemplos con dichos métodos.

## 14.5 Ejemplos ilustrativos

**EJEMPLO 14.1**
*Comisiones por consultoría de un fondo mutualista*

Consulte los datos de la tabla 14.1 y el MRNL (14.2.2). Mediante la rutina de regresión no lineal del paquete *EViews 6*, que se vale del método de linealización,[7] obtuvimos los siguientes resultados de la regresión; los coeficientes, sus errores estándar y sus valores *t* se dan en *forma tabular*:

| Variable | Coeficiente | Error estándar | Valor *t* | Valor *p* |
|---|---|---|---|---|
| Intercepto | 0.5089 | 0.0074 | 68.2246 | 0.0000 |
| Activo | –0.0059 | 0.00048 | –12.3150 | 0.0000 |

$$R^2 = 0.9385 \qquad d = 0.3493$$

A partir de estos resultados expresamos el modelo estimado como:

$$\widehat{\text{Comisión}}_i = 0.5089\,\text{Activo}^{-0.0059} \tag{14.5.1}$$

Antes de analizar estos resultados hay que notar que si no se proporcionan los valores iniciales de los parámetros a fin de empezar el proceso de linealización, *EViews* lo hará. Le bastan cinco iteraciones a *EViews* para obtener los resultados mostrados en (14.5.1). No obstante, uno puede proporcionar sus propios valores iniciales para iniciar el proceso. A fin de demostrarlo, elegimos los valores iniciales para $\beta_1 = 0.45$ y $\beta_2 = 0.01$.Obtuvimos los mismos resultados que en (14.5.1), pero necesitamos ocho iteraciones. *Es importante mencionar que se requerirán menos iteraciones si los valores iniciales no están muy distantes de los valores finales*. En algunos casos pueden elegirse los valores iniciales de los parámetros tan sólo con una regresión por MCO de la regresada sobre la(s) regresadora(s), únicamente ignorando la no linealidad. Por ejemplo, con los datos de la tabla 14.1, si fuese a efectuar la regresión de las comisiones sobre los activos, la estimación por MCO de $\beta_1$ sería 0.5028 y el de $\beta_2$ sería de –0.002, los cuales están mucho más

[6] Hay otro método que a veces se utiliza —**método Marquard**— y que es un procedimiento intermedio entre el método de pasos descendente y el de linealización (o series de Taylor). El lector interesado puede consultar las referencias para conocer mayores detalles de dicho método.

[7] *EViews* proporciona tres métodos: el de la ruta ascendente cuadrática, el de Newton-Raphson y el de Berndt-Hall-Hall-Hausman. La opción por omisión es la ruta ascendente cuadrática, que es una variación del método de Newton-Raphson.

**EJEMPLO 14.1** (*continuación*)

cerca de los valores finales dados en (14.5.1). (Para los detalles técnicos consulte el apéndice 14A, sección 14A.3.)

Ahora bien, respecto de las propiedades de los estimadores de MCNL, recuerde que en el caso de los modelos de regresión lineales con términos de error distribuidos normalmente, pudimos desarrollar procesos de inferencia exactos (es decir, pruebas de hipótesis) con las pruebas $t$, $F$ y ji cuadrada en muestras pequeñas y grandes. Por desgracia, no es el caso con los MRNL, aunque tengan términos de error distribuidos normalmente. *Los estimadores de MCNL no están distribuidos normalmente, no son insesgados y no tienen varianza mínima en muestras pequeñas o finitas.* Como resultado, no sirve la prueba $t$ (para probar la significancia de un coeficiente individual) ni la $F$ (para probar la significancia global de la regresión estimada), pues no puede obtenerse una estimación insesgada de la varianza del error $\sigma^2$ a partir de los residuos estimados. Es más, los residuos (la diferencia entre los valores $Y$ reales y los valores $Y$ estimados obtenidos del MRNL) no necesariamente suman cero. La suma de SCE y SCR no necesariamente da como resultado la SCT; por tanto, $R^2 =$ SCE/SCT puede no ser un estadístico descriptivo importante para tales modelos. No obstante, podemos calcular $R^2$ como:

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} \tag{14.5.2}$$

donde $Y =$ regresada y $\hat{u}_i = Y_i - \hat{Y}_i$, donde $\hat{Y}_i$ son los valores $Y$ estimados del MRNL (ajustado).

En consecuencia, las inferencias respecto de los parámetros de la regresión en una regresión no lineal suelen basarse en la teoría de las muestras grandes, según la cual, cuando la muestra es grande, los estimadores de mínimos cuadrados y de máxima verosimilitud para modelos de regresión no lineal con términos de error normales están casi normalmente distribuidos y casi son insesgados, además de que casi tienen varianza mínima. Esta teoría de muestras grandes también aplica cuando los términos de error no están normalmente distribuidos.[8]

Así, en resumen, todos los procedimientos de inferencia en los MRNL son de muestras grandes o asintóticos. Al retomar el ejemplo 14.1, los estadísticos $t$ de (14.5.1) tienen significado sólo si se interpretan en el contexto de muestras grandes. En ese sentido, podemos decir que los coeficientes estimados mostrados en (14.5.1) son estadísticamente significativos en forma individual. Por supuesto, la muestra en el presente ejemplo es más bien pequeña.

De nuevo con (14.5.1), ¿cómo calculamos la tasa de cambio de $Y$ (= comisión) respecto de $X$ (el tamaño de los activos)? Con las reglas básicas de las derivadas, el lector puede ver que la tasa de cambio de $Y$ respecto de $X$ es:

$$\frac{dY}{dX} = \beta_1 \beta_2 e^{\beta_2 X} = (-0.0059)(0.5089)e^{-0.0059X} \tag{14.5.3}$$

Como se observa, la tasa de cambio de la comisión depende del valor de los activos. Por ejemplo, si $X = 20$ (millones), la tasa esperada de cambio en las comisiones cobradas se obtiene de (14.5.3), la cual resulta casi del –0.0031%. Por supuesto, esta respuesta cambia en función del valor $X$ con que se calcule. Si se juzga con la $R^2$ como se calculó en (14.5.2), el valor $R^2$ igual a 0.9385 sugiere que el MRNL elegido se ajusta muy bien a los datos de la tabla 14.1. El valor Durbin-Watson estimado (igual a 0.3493) puede sugerir una autocorrelación o tal vez un error de especificación del modelo. Aunque existen procedimientos que se encargan de estos problemas, así como de la heteroscedasticidad en MRNL, no examinaremos esos temas aquí. El lector interesado puede consultar las referencias.

[8] John Neter, Michael H. Kutner, Christopher J. Nachtsheim y William Wasserman, *Applied Regression Analysis*, 3a. ed., Irwin, pp. 548-549.

**EJEMPLO 14.2**
*Función de producción Cobb-Douglas para la economía mexicana*

Consulte los datos proporcionados en el ejercicio 14.9 (tabla 14.3), referentes a la economía mexicana de 1955 a 1974. Verá que el MRNL dado en (14.1.4) se ajusta a los datos, en donde $Y$ = producción, $X_2$ = insumo trabajo y $X_3$ = insumo capital. Con *EViews 6* obtuvimos los siguientes resultados de regresión, después de 32 iteraciones.

| Variable | Coeficiente | Error estándar | Valor *t* | Valor *p* |
|---|---|---|---|---|
| Intercepto | 0.5292 | 0.2712 | 1.9511 | 0.0677 |
| Trabajo | 0.1810 | 0.1412 | 1.2814 | 0.2173 |
| Capital | 0.8827 | 0.0708 | 12.4658 | 0.0000 |

$$R^2 = 0.9942 \qquad d = 0.2899$$

Por tanto, la función de producción Cobb-Douglas estimada es:

$$\widehat{\text{PIB}}_t = 0.5292 \text{ Trabajo}_t^{0.1810} \text{ Capital}_t^{0.8827} \tag{14.5.4}$$

Interpretada de manera asintótica, la ecuación muestra que en el modelo sólo el coeficiente del insumo capital es significativo. En el ejercicio 14.9 se pide al lector comparar estos resultados con los obtenidos en la función de producción Cobb-Douglas multiplicativa, como se dio en (14.1.2).

**EJEMPLO 14.3**
*Crecimiento de la población de Estados Unidos, 1970-2007*

La tabla del ejercicio 14.8 presenta datos sobre la población total de Estados Unidos de 1970 a 2007. A menudo se mide el crecimiento de algunas poblaciones (seres humanos, bacterias, etc.) con un **modelo logístico** del siguiente tipo:

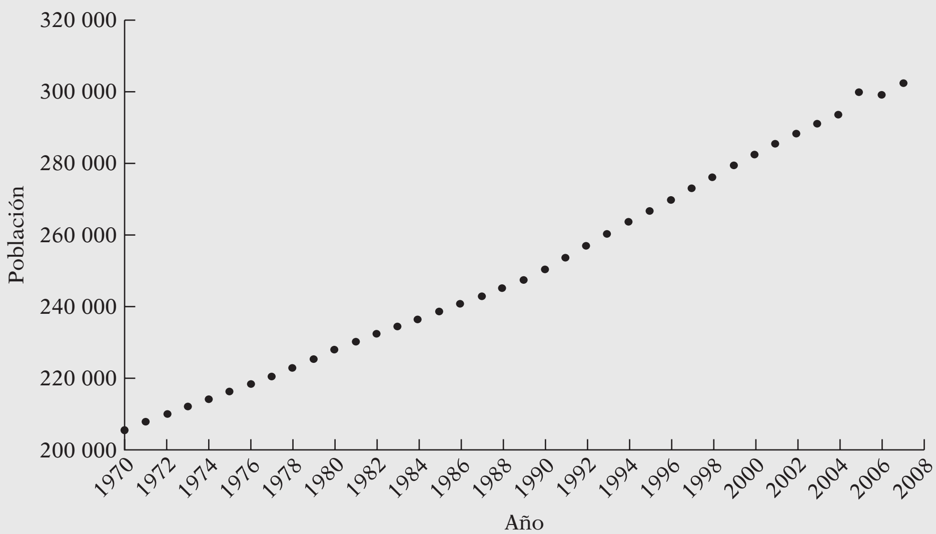
$$Y_t = \frac{\beta_1}{1 + e^{(\beta_2 + \beta_3^t)}} + u_t \tag{14.5.5}$$

donde $Y$ = población, en millones; $t$ = tiempo (medido cronológicamente), y las $\beta$ son los parámetros.

Este modelo es no lineal en los parámetros; no existe una manera sencilla de convertirlo en un modelo lineal en los parámetros. Por tanto, es necesario un método de estimación no lineal para estimar los parámetros. Observe una característica interesante de este modelo: a pesar de que sólo tiene dos variables, población y tiempo, hay tres parámetros desconocidos, lo cual muestra que en un MRNL puede haber más parámetros que variables.

Un intento por ajustar la ecuación (14.5.5) a los datos no tuvo éxito, porque todos los coeficientes estimados fueron estadísticamente insignificantes. Tal vez esto no deba sorprender, pues, si graficamos la población contra el tiempo, obtenemos la figura 14.2.

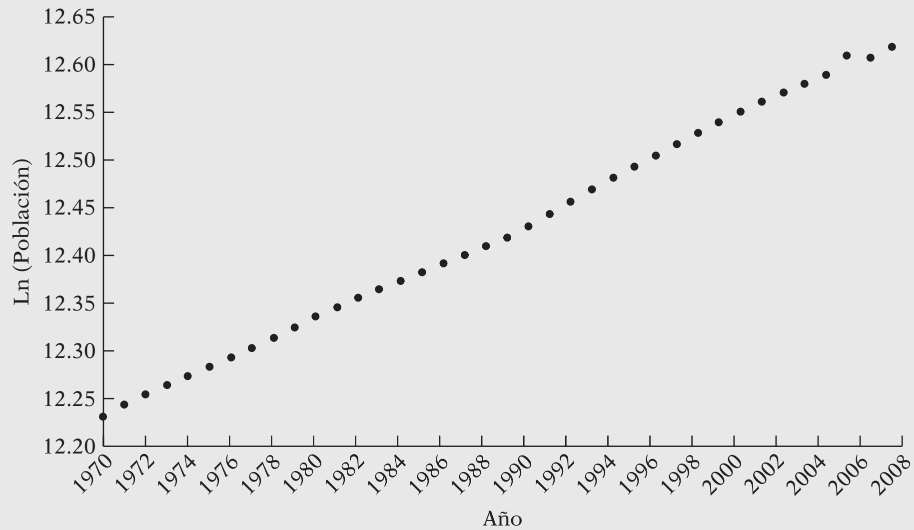
**FIGURA 14.2**
Población por año.



**EJEMPLO 14.3** (*continuación*)

Esta figura demuestra una relación casi lineal entre las dos variables. Si se graficamos el *logaritmo* de la población contra el tiempo, obtenemos la siguiente figura:

**FIGURA 14.3**
Logaritmo de población sobre el año.




La pendiente de esta figura (multiplicada por 100) da la tasa de crecimiento de la población (¿por qué?).

De hecho, si efectuamos una regresión del logaritmo de la población sobre el tiempo, obtenemos los siguientes resultados:

```
Variable dependiente: L Población
Método: Mínimos cuadrados
Muestra: 1970-2007
Observaciones incluidas: 38
```

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | −8.710413 | 0.147737 | −58.95892 | 0.0000 |
| AÑO | 0.010628 | 7.43E-05 | 143.0056 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.998244 | Media de la variable dependiente | 12.42405 |
| *R* cuadrada ajustada | 0.998195 | Desviación estándar de la variable dependiente | 0.118217 |
| Error estándar de la regresión | 0.005022 | Criterio de información de Akaike | −7.698713 |
| Suma de cuadrados residual | 0.000908 | Criterio de Schwarz | −7.612525 |
| Log verosimilitud | 148.2756 | Criterio de Hannan-Quinn | −7.668048 |
| Estadístico *F* | 20465.26 | Estadístico de Durbin-Watson | 0.366006 |
| Prob. (estadístico *F*) | 0.000000 | | |

Esta tabla muestra que, de 1970 a 2007, la población de Estados Unidos creció con una tasa aproximada de 1.06 por ciento anual. El valor $R^2$ de 0.998 revela un ajuste casi perfecto.

Este ejemplo pone de manifiesto un punto importante: a veces, un modelo lineal (en los parámetros) es preferible a uno no lineal (en los parámetros).

**EJEMPLO 14.4**
*Transformación Box-Cox: Población de Estados Unidos, 1970-2007*

En el apéndice 6A.5 consideramos brevemente la transformación Box-Cox. Seguiremos trabajando con el ejemplo 14.3, pero supondremos el siguiente modelo:

$$\text{Población}^{\lambda} = \beta_1 + \beta_2 \text{ Año} + u$$

Como se mencionó en el apéndice 6A.5, según el valor de λ tenemos las siguientes posibilidades:

| Valor de λ | Modelo |
|---|---|
| −1 | $\frac{1}{\text{Población}} = \beta_1 + \beta_2 \text{ Año} + u$ |
| 0 | $\ln \text{Población} = \beta_1 + \beta_2 \text{ Año} + u$ |
| 1 | $\text{Población}_i = \beta_1 + \beta_2 \text{ Año} + u$ |

El primero es un modelo inverso, el segundo es un modelo semilog (que ya estimamos en el ejemplo 14.3) y el tercero es un modelo lineal (en las variables).

¿Cuál de estos modelos es adecuado para los datos de población? Para responder se usa la rutina Box-Cox en STATA (versión 10):

| Prueba H0: | Log verosimilitud restringido | Estadístico RV ji$^2$ | Valor *p*, Prob. > ji$^2$ |
|---|---|---|---|
| $\theta = -1$ | −444.42475 | 0.14 | 0.707 |
| $\theta = 0$ | −444.38813 | 0.07 | 0.794 |
| $\theta = 1$ | −444.75684 | 0.81 | 0.369 |

*Nota:* En nuestra notación, theta ($\theta$) es lo mismo que lambda (λ). Esta tabla muestra que, con base en la prueba de la razón de verosimilitud (RV), no podemos rechazar ninguno de estos valores λ como posibles valores de la potencia de la población; es decir, en el presente ejemplo, los modelos lineal, inverso y semilog son candidatos por igual para representar el comportamiento de la población en el periodo de muestra de 1970 a 2007. Por tanto, presentamos los resultados de los tres modelos:

| Variable dependiente | Intercepto | Pendiente | $R^2$ |
|---|---|---|---|
| 1/Población | 0.000089 | −4.28e-08 | 0.9986 |
| | *t* (166.14) | (−1568.10) | |
| ln Población | −8.7104 | 0.0106 | 0.9982 |
| | *t* (−58.96) | (143.06) | |
| Población | −5042627 | 2661.825 | 0.9928 |
| | *t* (−66.92) | (70.24) | |

En todos estos modelos los coeficientes estimados son, sin excepción, muy significativos estadísticamente. Sin embargo, observe que los valores $R^2$ no son directamente comparables, porque las variables dependientes de los tres modelos son diferentes.

Este ejemplo demuestra que las técnicas de estimación no lineales son aplicables a situaciones concretas.

## Resumen y conclusiones

Los principales aspectos que se analizaron en este capítulo se resumen de la siguiente forma:

1. Aunque los modelos de regresión lineal predominan en la teoría y en la práctica, hay ocasiones en las que son útiles los modelos de regresión no lineales en los parámetros (MRNL).
2. Las matemáticas propias de los modelos de regresión lineal son comparablemente sencillas, pues ofrecen soluciones explícitas o analíticas para los coeficientes de tales modelos. La teoría de inferencia de muestras grandes y muestras pequeñas para dichos modelos está bien fundamentada.
3. En contraste, para los modelos de regresión intrínsecamente no lineales, los valores de los parámetros no se obtienen de manera explícita. Deben calcularse de forma numérica; es decir, mediante procesos iterativos.
4. Existen diversos métodos para obtener los estimados de los MRNL, como 1) ensayo y error, 2) mínimos cuadrados no lineales (MCNL) y 3) linealización mediante la expansión de series de Taylor.
5. Los paquetes de computación ahora cuentan con rutinas integradas, como las de Gauss-Newton, Newton-Raphson y la de Marquard. Todas son rutinas iterativas.
6. Los estimadores de MCNL no presentan propiedades óptimas en las muestras finitas, pero en muestras grandes sí cuentan con tales propiedades. En consecuencia, los resultados de los MCNL en pequeñas muestras deben interpretarse con precaución.
7. Los problemas de autocorrelación, heteroscedasticidad y especificación de modelos pueden afectar a los MRNL, como afectan a los modelos de regresión lineales.
8. Ilustramos los MCNL con diversos ejemplos. Gracias a la gran disponibilidad de paquetes de computación accesibles, la estimación de los MCNL ya no es un obstáculo. Por consiguiente, el lector no debe evitar estos modelos cuando sean convenientes por razones prácticas o teóricas. De hecho, si consulta el ejercicio 12.10, verá que en la ecuación (1) existe un modelo de regresión intrínsecamente no lineal, que debe calcularse como tal.

## EJERCICIOS

### Preguntas

14.1. ¿Qué se quiere decir con modelos de regresión intrínsecamente lineales e intrínsecamente no lineales? Proporcione algunos ejemplos.

14.2. Como el término de error en la función producción de Cobb-Douglas se introduce mediante una forma aditiva o una multiplicativa, ¿cómo decidiría entre ambas?

14.3. ¿Cuál es la diferencia entre la estimación por MCO y por mínimos cuadrados no lineales (MCNL)?

14.4. La relación entre la presión y la temperatura en el vapor saturado se expresa como:*

$$Y = \beta_1(10)^{\beta_2 t/(\gamma+t)} + u_t$$

donde $Y$ = presión y $t$ = temperatura. Con el método de los mínimos cuadrados no lineales (MCNL), obtenga las *ecuaciones normales* para este modelo.

* Adaptado de Draper y Smith, *op. cit.*, p. 554.

14.5. Establezca si los siguientes enunciados son verdaderos o falsos. Justifique sus respuestas.

*a*) La inferencia estadística en la regresión con MCNL no puede basarse en las pruebas $t$, $F$ y $\chi^2$ usuales, aunque se suponga que el término de error está normalmente distribuido.

*b*) El coeficiente de determinación ($R^2$) no es un número particularmente importante para un MRNL.

14.6. ¿Cómo linealizaría la función de producción ECS analizada en este capítulo? Muestre los pasos necesarios.

14.7. Los modelos que describen el comportamiento de una variable a lo largo del tiempo se llaman modelos de crecimiento. Tales modelos operan en diversos campos, como economía, biología, botánica, ecología y demografía. Los modelos de crecimiento adoptan varias formas, tanto lineales como no lineales. Considere los siguientes modelos, donde $Y$ es la variable cuyo crecimiento se desea medir, $t$ es el tiempo medido cronológicamente y $u_t$ es el término de error estocástico.

*a*) $Y_t = \beta_1 + \beta_2 t + u_t$

*b*) $\ln Y_t = \beta_1 + \beta_2 t + u_t$

*c*) Modelo de crecimiento logístico: $Y_t = \frac{\beta_1}{1 + \beta_2 e^{-\beta_3 t}} + u_t$

*d*) Modelo de crecimiento de Gompertz: $Y_t = \beta_1 e^{-\beta_2 e^{-\beta_3 t}} + u_t$

Determine las propiedades de estos modelos al considerar el crecimiento de $Y$ en relación con el tiempo.

## Ejercicios empíricos

14.8. Los datos de la tabla 14.2 corresponden a la población de Estados Unidos (en millones de personas) de 1970 a 2007. Ajuste los modelos de crecimiento que se presentan en el ejercicio 14.7 y elija el modelo que mejor se ajuste. Interprete los parámetros del modelo.

14.9. La tabla 14.3 presenta datos reales sobre el PIB, trabajo y capital de México de 1955 a 1974. Vea si la función de producción Cobb-Douglas multiplicativa, dada en la ecuación (14.1.2*a*), se ajusta a estos datos. Compare esos resultados con los obtenidos del ajuste de la función de producción Cobb-Douglas aditiva, dada en (14.1.4), cuyos resultados se presentan en el ejemplo 14.2. ¿Cuál ajusta mejor?

**TABLA 14.2**
**Población de Estados Unidos (millones de habitantes)**

Fuente: *Economic Report of the President*, 2008.

| Año | Población | Año | Población |
|---|---|---|---|
| 1970 | 205 052 | 1989 | 247 342 |
| 1971 | 207 661 | 1990 | 250 132 |
| 1972 | 209 896 | 1991 | 253 493 |
| 1973 | 211 909 | 1992 | 256 894 |
| 1974 | 213 854 | 1993 | 260 255 |
| 1975 | 215 973 | 1994 | 263 436 |
| 1976 | 218 035 | 1995 | 266 557 |
| 1977 | 220 239 | 1996 | 269 667 |
| 1978 | 222 585 | 1997 | 272 912 |
| 1979 | 225 055 | 1998 | 276 115 |
| 1980 | 227 726 | 1999 | 279 295 |
| 1981 | 229 966 | 2000 | 282 407 |
| 1982 | 232 188 | 2001 | 285 339 |
| 1983 | 234 307 | 2002 | 288 189 |
| 1984 | 236 348 | 2003 | 290 941 |
| 1985 | 238 466 | 2004 | 293 609 |
| 1986 | 240 651 | 2005 | 299 801 |
| 1987 | 242 804 | 2006 | 299 157 |
| 1988 | 245 021 | 2007 | 302 405 |

**TABLA 14.3** **Datos de la función de producción de la economía mexicana**

| Observación | PIB | Trabajo | Capital | Observación | PIB | Trabajo | Capital |
|---|---|---|---|---|---|---|---|
| 1955 | 114 043 | 8 310 | 182 113 | 1965 | 212 323 | 11 746 | 315 715 |
| 1956 | 120 410 | 8 529 | 193 749 | 1966 | 226 977 | 11 521 | 337 642 |
| 1957 | 129 187 | 8 738 | 205 192 | 1967 | 241 194 | 11 540 | 363 599 |
| 1958 | 134 705 | 8 952 | 215 130 | 1968 | 260 881 | 12 066 | 391 847 |
| 1959 | 139 960 | 9 171 | 225 021 | 1969 | 277 498 | 12 297 | 422 382 |
| 1960 | 150 511 | 9 569 | 237 026 | 1970 | 296 530 | 12 955 | 455 049 |
| 1961 | 157 897 | 9 527 | 248 897 | 1971 | 306 712 | 13 338 | 484 677 |
| 1962 | 165 286 | 9 662 | 260 661 | 1972 | 329 030 | 13 738 | 520 553 |
| 1963 | 178 491 | 10 334 | 275 466 | 1973 | 354 057 | 15 924 | 561 531 |
| 1964 | 199 457 | 10 981 | 295 378 | 1974 | 374 977 | 14 154 | 609 825 |

*Notas:* El PIB se expresa en millones de pesos de 1960.
El trabajo se expresa en miles de personas.
El capital se expresa en millones de pesos de 1960.
Fuente: Víctor J. Elías, *Sources of Growth: A Study of Seven Latin American Economies*, International Center for Economic Growth, ICS Press, San Francisco, 1992, tablas E-5, E-12 y E-14.

# Apéndice 14A

## 14A.1 Derivación de las ecuaciones (14.2.4) y (14.2.5)

Exprese (14.2.2) como

$$u_i = Y_i - \beta_1 e^{\beta_2 X_i} \tag{1}$$

Por consiguiente,

$$\sum u_i^2 = \sum \left(Y_i - \beta_1 e^{\beta_2 X_i}\right)^2 \tag{2}$$

La suma de cuadrados del error es por tanto una función de $\beta_1$ y $\beta_2$, pues se conocen los valores de $Y$ y $X$. En consecuencia, para reducir la suma de cuadrados del error, se diferencia parcialmente respecto de dos incógnitas, lo cual da:

$$\frac{\partial \sum u_i^2}{\partial \beta_1} = 2\sum \left(Y_i - \beta_1 e^{\beta_2 X_i}\right)\left(-1 e^{\beta_2 X_i}\right) \tag{3}$$

$$\frac{\partial \sum u_i^2}{\partial \beta_2} = 2\sum \left(Y_i - \beta_1 e^{\beta_2 X_i}\right)\left(-\beta_1 e^{\beta_2 X_i} X_i\right) \tag{4}$$

Debido a la condición de optimización de primer orden, al igualar las ecuaciones anteriores a cero y resolverlas simultáneamente, obtenemos (14.2.4) y (14.2.5). Observe que diferenciamos la suma de cuadrados de los errores con la regla de la cadena.

## 14A.2 Método de linealización

Los estudiantes familiarizados con el cálculo recordarán el **teorema de Taylor**, el cual establece que toda función arbitraria $f(X)$ que sea continua y tenga derivadas de orden $n$ continuas puede aproximarse alrededor del punto $X = X_0$ mediante una función polinomial y un residuo de la siguiente forma:

$$f(X) = \frac{f(X_0)}{0!} + \frac{f'(X_0)(X - X_0)}{1!} + \frac{f''(X_0)(X - X_0)^2}{2!} + \cdots$$

$$+ \frac{f^n(X_0)(X - X_0)^n}{n!} + R \tag{1}$$

donde $f'(X_0)$ es la primera derivada de $f(X)$ evaluada en $X = X_0$, $f''(X_0)$ es la segunda derivada de $f(X)$ evaluada en $X = X_0$, y así sucesivamente; además, $n!$ (se lee como $n$ factorial) significa $n(n-1)(n-2)\ldots 1$, con la convención de que $0! = 1$, y $R$ representa al residuo. Si tomamos $n = 1$, obtenemos una aproximación lineal; al elegir $n = 2$, obtenemos una aproximación polinomial de segundo grado. Como se esperaría, mientras mayor sea el orden del polinomio, mejor será la aproximación a la función original. Las series en (1) se conocen como la **expansión de series de Taylor de $f(X)$ alrededor del punto $X = X_0$**. Como ejemplo, considere la función:

$$Y = f(X) = \alpha_1 + \alpha_2 X + \alpha_3 X^2 + \alpha_4 X^3$$

Suponga que deseamos aproximarla a $X = 0$. Ahora obtenemos:

$$f(0) = \alpha_1 \qquad f'(0) = \alpha_2 \qquad f''(0) = 2\alpha_3 \qquad f'''(0) = 6\alpha_4$$

Por consiguiente, obtenemos las siguientes aproximaciones:

Primer orden: $$Y = \alpha_1 + \frac{f'(0)}{1!} = \alpha_1 + a_2 X + \text{a residuo}\left(= \alpha_3 X^2 + \alpha_4 X^3\right)$$

Segundo orden: $$Y = f(0) + \frac{f'(0)}{1!}X + \frac{f''(0)}{2!}X^2$$
$$= \alpha_1 + \alpha_2 X + \alpha_3 X^2 + \text{a residuo}\left(= \alpha_4 X^3\right)$$

Tercer orden: $$Y = \alpha_1 + \alpha_2 X + \alpha_3 X^2 + \alpha_4 X^3$$

La aproximación de tercer orden reproduce con exactitud la ecuación original.

El objetivo de la aproximación mediante las series de Taylor con frecuencia consiste en elegir un polinomio de orden inferior, con la esperanza de que el término del residuo no tenga consecuencia alguna para el proceso. Suele utilizarse para aproximar una función no lineal mediante una función lineal, al eliminarse los términos de orden superior.

La aproximación mediante series de Taylor se generaliza con facilidad a una función con más de una $X$. Por ejemplo, considere la siguiente:

$$Y = f(X, Z) \tag{2}$$

y suponga que deseamos expandir alrededor de $X = a$ y $Z = b$. El teorema de Taylor demuestra que

$$\begin{aligned} f(x, z) &= f(a, b) + f_x(a, b)(x - a) \\ &\quad + f_z(a, b) f(z - b) + \frac{1}{2!}[f_{xx}(a, b)(x - a)^2 \\ &\quad - 2 f_{xz}(a, b)(x - a)(z - b) + f_{zz}(a, b)(z - b)^2] + \cdots \end{aligned} \tag{3}$$

donde $f_x$ = la derivada parcial de la función respecto de $X$, $f_{xx}$ = la segunda derivada parcial respecto de $X$ y de manera semejante para la variable $Z$. Si queremos una aproximación lineal para la función, utilizamos los dos primeros términos de (3); si queremos una aproximación cuadrática, o de segundo orden, empleamos los tres primeros términos de (3), y así sucesivamente.

## 14A.3 Aproximación lineal de la función exponencial dada en (14.2.2)

La función en consideración es:

$$Y = f(\beta_1, \beta_2) = \beta_1 e^{\beta_2 X} \tag{1}$$

*Nota:* Para un manejo sencillo, eliminamos los subíndices.

Recuerde que en esta función las incógnitas son los coeficientes $\beta$. Linealicemos esta función en $\beta_1 = \beta_1^*$ y $\beta_2 = \beta_2^*$, donde las cantidades con asterisco son *valores fijos dados*. Para *linealizar* lo anterior, procedemos de la siguiente forma:

$$Y = f(\beta_1, \beta_2) = f(\beta_1^*, \beta_2^*) + f_{\beta_1}(\beta_1^*, \beta_2^*)(\beta_1 - \beta_1^*) + f_{\beta_2}(\beta_1^*, \beta_2^*)(\beta_2 - \beta_2^*) \tag{2}$$

donde $f_{\beta 1}$ y $f_{\beta 2}$ son las derivadas parciales de la función (1) respecto de las incógnitas, y tales derivadas se evaluarán en los valores (supuestos) con asterisco de los parámetros desconocidos. Observe que sólo usamos las primeras derivadas en la expresión anterior, pues estamos linealizando la función. Ahora suponga que $\beta_1^* = 0.45$ y $\beta_2^* = 0.01$, las cuales son sólo conjeturas de los verdaderos coeficientes. Ahora

$$f(\beta_1^* = 0.45, \beta_2^* = 0.01) = 0.45e^{0.01X_i}$$
$$f_{\beta_1} = e^{\beta_2 X_i} \quad \text{y} \quad f_{\beta_2} = \beta_1 X_i e^{\beta_2 X_i} \tag{3}$$

debido a las reglas usuales de la derivación. Al evaluar las derivadas anteriores en los valores dados y al regresar a (2), obtenemos:

$$Y_i = 0.45e^{0.01X_i} + e^{0.01X_i}(\beta_1 - 0.45) + (0.45)X_i e^{0.01X_i}(\beta_2 - 0.01) \tag{4}$$

que escribimos como:

$$(Y_i - 0.45e^{0.01X_i}) = e^{0.01X_i}\alpha_1 + 0.45X_i e^{0.01X_i}\alpha_2 \tag{5}$$

donde

$$\alpha_1 = (\beta_1 - 0.45) \quad \text{y} \quad \alpha_2 = (\beta_2 - 0.01) \tag{6}$$

Ahora, sea $Y_i^* = (Y_i - 0.45e^{0.01X_i})$, $X_1 = e^{0.01X_i}$ y $X_{2i} = 0.45X_i e^{0.01X_i}$. Con estas definiciones y al sumar el término error $u_i$, podemos finalmente escribir (5) como:

$$Y_i^* = \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_i \tag{7}$$

Y aquí tenemos un modelo de regresión lineal. Como $Y_i^*$, $X_{1i}$ y $X_{2i}$ se calculan con facilidad a partir de los datos, podemos estimar con sencillez (7) mediante MCO y obtener los valores de $\alpha_1$ y $\alpha_2$. Así, a partir de (6) obtenemos:

$$\beta_1 = \hat{\alpha}_1 + 0.45 \quad \text{y} \quad \beta_2 = \hat{\alpha}_2 + 0.01 \tag{8}$$

Llame a esos valores $\beta_1^{**}$ y $\beta_2^{**}$, respectivamente. Con esos valores (revisados) podemos empezar el proceso iterativo dado en (2), con lo cual obtenemos otro conjunto más de valores de los coeficientes $\beta$. Podemos seguir iterando (o linealizando) de esta forma hasta que no haya cambios sustanciales en los valores de los coeficientes $\beta$. En el ejemplo 14.1 se requirieron cinco iteraciones, pero para el ejemplo de la función Cobb-Douglas de México (ejemplo 14.2) se necesitaron 32 iteraciones. No obstante, la lógica de estas iteraciones es el procedimiento que acabamos de ejemplificar.

Para la estructura de las comisiones del fondo mutualista de la sección 14.3, las $Y^*$, $X_1$ y $X_2$ se presentan en la tabla 14.4 tal como aparecen en (6); los datos básicos se proporcionan en la tabla 14.1. Con base en estos valores, los resultados de la regresión correspondientes a (7) son:

Variable dependiente: Y*
Método: Mínimos cuadrados

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| $X_1$ | 0.022739 | 0.014126 | 1.609705 | 0.1385 |
| $X_2$ | -0.010693 | 0.000790 | -13.52900 | 0.0000 |

$R^2$ = 0.968324 Estadístico *d* de Durbin-Watson = 0.308883

Ahora, con (8), el lector puede verificar que

$$\beta_1^* = 0.4727 \quad \text{y} \quad \beta_2^* = -0.00069 \tag{9}$$

**TABLA 14.4**

| $Y^*$ | $X_1$ | $X_2$ |
|---|---|---|
| 0.067744 | 1.005013 | 0.226128 |
| 0.034928 | 1.051271 | 2.365360 |
| −0.013327 | 1.105171 | 4.973269 |
| −0.062825 | 1.161834 | 7.842381 |
| −0.109831 | 1.221403 | 10.99262 |
| −0.154011 | 1.284025 | 14.44529 |
| −0.195936 | 1.349859 | 18.22309 |
| −0.236580 | 1.419068 | 22.35031 |
| −0.276921 | 1.491825 | 26.85284 |
| −0.317740 | 1.568312 | 31.75832 |
| −0.397464 | 1.733253 | 42.89801 |
| −0.446153 | 1.822119 | 49.19721 |

Compare estos números con las conjeturas iniciales de 0.45 y 0.01, respectivamente, para los dos parámetros. Con las nuevas estimaciones dadas en (9) podemos comenzar el procedimiento iterativo una vez más y seguirlo hasta que haya "convergencia" en el sentido de que la rutina final de estimaciones no difiera mucho de la rutina anterior. Por supuesto, necesitará menos iteraciones si la conjetura inicial está más próxima a los valores finales. Asimismo, observe que sólo utilizamos el término lineal de la expansión de series de Taylor. Si fuese a utilizar términos cuadráticos o superiores en la expansión, tal vez alcance los valores finales con mayor rapidez. Pero en muchas aplicaciones la aproximación lineal ha demostrado ser muy adecuada.

# Capítulo 15

# Modelos de regresión de respuesta cualitativa

En todos los modelos de regresión considerados hasta ahora, supusimos que la regresada, la variable dependiente o la variable de *respuesta Y* era cuantitativa, mientras que las variables explicativas podían ser cuantitativas o cualitativas, o una mezcla de las dos. De hecho, en el capítulo 9, sobre variables dicótomas, vimos cómo se introducen las regresoras dicótomas en un modelo de regresión y el papel que desempeñan en situaciones específicas.

En este capítulo consideraremos modelos de regresión en los cuales la variable dependiente o de respuesta puede ser en sí misma de naturaleza cualitativa. Aunque los modelos de regresión de respuesta cualitativa cada vez son más comunes en diversas áreas de las ciencias sociales y la investigación médica, plantean interesantes retos respecto de su cálculo y estimación. En este capítulo sólo abordaremos algunos temas importantes de esta área, y dejamos los detalles para libros más especializados.[1]

## 15.1 Naturaleza de los modelos de respuesta cualitativa

Suponga que deseamos estudiar la participación en la fuerza laboral (PFL) de los hombres adultos. Como un adulto está en la fuerza laboral o no lo está, la PFL es una decisión de *sí* o *no.* Por consiguiente, la variable de respuesta, o regresada, sólo puede adquirir dos valores; por ejemplo, 1 si la persona está en la fuerza laboral y 0 si no lo está. En otras palabras, la regresada es una **variable binaria**, o **dicótoma**. La investigación en economía del trabajo indica que la PFL es una función de la tasa de desempleo, tasa del salario promedio, escolaridad, ingreso familiar, etcétera.

Otro ejemplo: piense en las elecciones presidenciales de Estados Unidos. Suponga que hay dos partidos políticos, el Demócrata y el Republicano. La variable dependiente en este caso es la elección de voto entre ambos partidos políticos. Suponga que $Y = 1$ si el voto es para el candidato demócrata y $Y = 0$ si el voto es republicano. El economista Ray Fair, de la Universidad Yale, y otros científicos de la política han efectuado una gran cantidad de trabajo de investigación sobre este tema.[2] Algunas variables en la elección del voto son la tasa del crecimiento del PIB, las tasas

[1] En un nivel introductorio, el lector puede encontrar muy útiles las siguientes fuentes: Daniel A. Powers y Yu Xie, *Statistical Methods for Categorical Data Analysis,* Academic Press, 2000; John H. Aldrich y Forrest Nelson, *Linear Probability, Logit and Probit Models,* Sage, 1984; y Tim Futing Liao, *Interpreting Probability Models: Logit, Probit and Other Generalized Linear Models,* Sage, 1994. Para una revisión muy completa de la bibliografía, véase G.S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics,* Cambridge University Press, 1983.

[2] Véase, por ejemplo, Ray Fair, "Econometrics and Presidential Elections", *Journal of Economic Perspective*, verano, 1996, pp. 89-102, y Machael S. Lewis-Beck, *Economics and Elections: The Major Western Democracies*, University of Michigan Press, Ann Arbor, 1980.

de desempleo e inflación, si el candidato se va a reelegir, etc. Para nuestros propósitos, lo importante es notar que la regresada es una variable cualitativa.

Podemos pensar en otros ejemplos en los que la regresada es cualitativa por naturaleza. Por consiguiente, una familia posee o no posee casa, tiene seguro contra invalidez o no lo tiene, ambos cónyuges están en la fuerza laboral o sólo uno de ellos lo está. En forma similar, un determinado fármaco es eficaz para curar una enfermedad o no lo es. Una empresa decide declarar el rendimiento de sus acciones u ocultarlo, un senador decide votar en favor de un impuesto o en contra, el presidente decide vetar una ley o aprobarla, etcétera.

No tenemos que restringir la variable de respuesta a un sí o no, o sólo a categorías dicótomas. De regreso con el ejemplo de las elecciones presidenciales, suponga que existen tres partidos: el Demócrata, el Republicano y el Independiente. La variable de respuesta en este caso es **tricótoma**. En general, podemos tener una variable de respuesta **policótoma** (o de **categoría múltiple**).

En primer lugar, lo que pretendemos es considerar la regresada dicótoma y luego analizar diversas extensiones del modelo básico. Pero antes, es importante hacer hincapié en una diferencia fundamental entre un modelo de regresión en el que la regresada $Y$ es cuantitativa y un modelo en el que es cualitativa.

En un modelo en donde $Y$ es cuantitativa, el objetivo consiste en estimar su valor esperado, o media esperada, dados los valores de las regresoras. En términos del capítulo 2, lo que deseamos es obtener $E(Y_i \mid X_{1i}, X_{2i}, \ldots, X_{ki})$, donde las $X$ son las regresoras, tanto cuantitativas como cualitativas. En los modelos en donde $Y$ es cualitativa, el objetivo es encontrar la probabilidad de que un acontecimiento suceda, como votar por el candidato demócrata, poseer una casa, pertenecer a un sindicato, practicar algún deporte, etc. Por tanto, los modelos de regresión con respuestas cualitativas a menudo se conocen como *modelos de probabilidad*.

En lo que resta de este capítulo buscaremos respuestas a las siguientes preguntas:

1. ¿Cómo se estiman los modelos de regresión con respuestas cualitativas?, ¿simplemente se estiman con los procedimientos usuales de MCO?
2. ¿Se presentan problemas de inferencia especiales? En otras palabras, ¿el procedimiento de pruebas de hipótesis se diferencia de los que hemos estudiado hasta este momento?
3. Si una regresada es cualitativa, ¿cómo se mide la bondad de ajuste de dichos modelos?, ¿tiene algún valor la $R^2$ convencionalmente calculada para tales modelos?
4. Una vez rebasado el caso de la regresada dicótoma, ¿cómo se estiman e interpretan los modelos de regresión policótomos? Asimismo, ¿cómo se trabaja con modelos en los que la regresada es **ordinal**; es decir, consiste en una variable categórica ordenada, como la escolaridad (menos de ocho años, de ocho a 11 años, 12 años y 13 años o más), o la regresada es **nominal**, en la que no existe un orden inherente, como el origen étnico (blanco, negro, hispano, asiático, etcétera)?
5. ¿Cómo se elaboran los modelos para fenómenos como el número anual de visitas al médico, la cantidad de patentes que registra una empresa en un año determinado, el número de artículos publicados por un profesor universitario durante un año, el número de llamadas telefónicas recibidas en un lapso de cinco minutos o la cantidad de automóviles que pasan por una caseta de cobro durante cinco minutos? Dichos fenómenos, llamados **datos de conteo**, o **sucesos raros**, son un ejemplo del proceso (de probabilidad) de **Poisson**.

En este capítulo responderemos varias de estas interrogantes en un nivel elemental, pues algunos temas son de carácter muy elevado y requieren un conocimiento matemático y estadístico mayor que el supuesto para este libro. Consulte la bibliografía en las notas para conocer con mayor detalle la materia.

Comenzaremos el estudio de los modelos con respuesta cualitativa, en primer lugar, el modelo de regresión con **respuesta binaria**. Hay cuatro métodos para crear un modelo de probabilidad para una variable de respuesta binaria:

1. El **modelo lineal de probabilidad (MLP)**

2. El **modelo logit**
3. El **modelo probit**
4. El **modelo tobit**

En vista de su simplicidad relativa y debido a que se estima mediante mínimos cuadrados ordinarios (MCO), estudiaremos primero el MLP, y dejamos los otros tres modelos para las secciones subsiguientes.

## 15.2 Modelo lineal de probabilidad (MLP)

Para establecer las ideas, considere el siguiente modelo simple:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{15.2.1}$$

donde $X$ = el ingreso familiar, y $Y = 1$ si la familia tiene casa propia y 0 si la familia no tiene casa propia.

El modelo (15.2.1) parece un modelo de regresión lineal común, pero debido a que la variable regresada es binaria, o dicótoma, se denomina **modelo lineal de probabilidad (MLP)**. Esto es porque la expectativa condicional de $Y_i$ dado $X_i$, $E(Y_i \mid X_i)$ puede interpretarse como la *probabilidad condicional* de que el suceso tenga lugar dado $X_i$; es decir, Pr $(Y_i = 1 \mid X_i)$. Así, en el ejemplo, $E(Y_i \mid X_i)$ da la probabilidad de que una familia tenga casa propia y perciba ingresos por una cierta cantidad $X_i$.

La justificación del nombre MLP para modelos como (15.2.1) es la siguiente: En el supuesto de que $E(u_i) = 0$, como de costumbre (para obtener estimadores insesgados), obtenemos

$$E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i \tag{15.2.2}$$

Ahora, si $P_i$ = probabilidad de que $Y_i = 1$ (es decir, de que el suceso ocurra) y $(1 - P_i)$ = probabilidad de que $Y_i = 0$ (es decir, de que el suceso no ocurra), la variable $Y_i$ tiene la siguiente distribución (de probabilidad):

| $Y_i$ | Probabilidad |
|---|---|
| 0 | $1 - P_i$ |
| 1 | $P_i$ |
| Total | 1 |

Es decir, $Y_i$ sigue la **distribución de probabilidades de Bernoulli**.

Por consiguiente, por la definición de esperanza matemática, obtenemos

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \tag{15.2.3}$$

Al comparar (15.2.2) con (15.2.3), igualamos

$$E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i = P_i \tag{15.2.4}$$

es decir, la esperanza condicional del modelo (15.2.1) en realidad se interpreta como la probabilidad condicional de $Y_i$. En general, la esperanza de una variable aleatoria Bernoulli está dada por la probabilidad de que esa variable sea igual a 1. Por cierto, observe que si existen $n$ intentos independientes, cada uno con una probabilidad $p$ de éxito y una probabilidad $(1 - p)$ de fracaso, y $X$ de tales intentos representa el número de éxitos, se dice que $X$ sigue una **distribución binomial**. La media de la distribución binomial es $np$, y su varianza, $np(1 - p)$. El término *éxito* se define dentro del contexto del problema.

Como la probabilidad $P_i$ debe encontrarse entre 0 y 1, tenemos la restricción

$$0 \leq E(Y_i \mid X_i) \leq 1 \tag{15.2.5}$$

es decir, la esperanza condicional (o probabilidad condicional) debe encontrarse entre 0 y 1.

Del análisis anterior parecería que los MCO pueden extenderse sin dificultad a modelos de regresión con variable dependiente binaria. Así, tal vez no haya ninguna novedad aquí. Por desgracia, no es el caso, pues el MLP plantea diversos problemas, a saber:

## No normalidad de las perturbaciones $u_i$

Aunque MCO no requiere que las perturbaciones (las $u_i$) estén normalmente distribuidas, supusimos que lo estaban para fines de inferencia estadística.[3] Pero el supuesto de normalidad para $u_i$ ya no se mantiene en los MLP porque, al igual que $Y_i$, $u_i$ sólo toma dos valores; es decir, también sigue la distribución de Bernoulli. Para ver esto, escribimos (15.2.1) como

$$u_i = Y_i - \beta_1 - \beta_2 X_i \tag{15.2.6}$$

La distribución de probabilidades de $u_i$ es

| | $u_i$ | Probabilidad |
|---|---|---|
| Cuando $Y_i = 1$ | $1 - \beta_1 - \beta_2 X_i$ | $P_i$ |
| Cuando $Y_i = 0$ | $-\beta_1 - \beta_2 X_i$ | $(1 - P_i)$ |

(15.2.7)

Obvio, no puede suponerse que $u_i$ esté normalmente distribuida; en realidad sigue la distribución de Bernoulli.

Pero el no cumplimiento del supuesto de normalidad quizá no sea tan crítico como parece porque sabemos que las estimaciones puntuales de MCO aún permanecen insesgadas (recuerde que si el objetivo es la estimación puntual, el supuesto de normalidad resulta innecesario). Además, puede demostrarse que, conforme el tamaño de la muestra aumenta indefinidamente, los estimadores de MCO tienden a tener una distribución[4] normal, según la teoría estadística. Por consiguiente, en muestras grandes, la inferencia estadística del MLP seguirá el procedimiento de MCO usual según el supuesto de normalidad.

## Varianzas heteroscedásticas de las perturbaciones

Aunque $E(u_i) = 0$ y cov $(u_i, u_j) = 0$ para $i \neq j$ (es decir, no hay correlación serial), ya no es posible sostener la afirmación de que las perturbaciones en el MLP son homoscedásticas. Esto, sin embargo, no debe sorprender. Como demuestra la teoría estadística, para una distribución de Bernoulli, la media y la varianza teóricas son $p$ y $p(1 - p)$ respectivamente, donde $p$ es la probabilidad de éxito (es decir, de que algo suceda), lo cual revela que la varianza es una función de la media. Por tanto, la varianza del error es heteroscedástica.

Para la distribución del término de error dado en (15.2.7), si se aplica la definición de varianza, el lector verificará que (véase el ejercicio 15.10)

$$\text{var}\,(u_i) = P_i(1 - P_i) \tag{15.2.8}$$

[3] Recuerde que recomendamos verificar el supuesto de normalidad mediante pruebas de normalidad adecuadas, como la de Jarque–Bera.

[4] La prueba se basa en el teorema central del límite y se encuentra en E. Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1966, pp. 195-197. Si las regresoras se consideran estocásticas y están normalmente distribuidas en forma conjunta, las pruebas *F* y *t* son aún útiles aunque las perturbaciones no sean normales. También considere que, conforme el tamaño de la muestra se incrementa de manera indefinida, la distribución binomial converge a la distribución normal.

Es decir, la varianza del término de error en el MLP es heteroscedástica. Como $P_i = E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$, la varianza de $u_i$ depende, al final, de los valores de $X$ y por tanto no es homoscedástica.

Ya sabemos que, en presencia de heteroscedasticidad, los estimadores de MCO, aunque insesgados, no son eficientes; es decir, no tienen varianza mínima. Pero, de nuevo, el problema de heteroscedasticidad, al igual que el de la no normalidad, no es insuperable. En el capítulo 11 analizamos diversos métodos para tratar el problema de heteroscedasticidad. Como la varianza de $u_i$ depende de $E(Y_i \mid X_i)$, una forma de resolver el problema de heteroscedasticidad es transformar el modelo (15.2.1) dividiendo ambos lados del modelo (15.2.1) entre

$$\sqrt{E(Y_i|X_i)[1 - E(Y_i|X_i)]} = \sqrt{P_i(1 - P_i)} = \text{digamos } \sqrt{w_i}$$

o sea

$$\frac{Y_i}{\sqrt{w_i}} = \frac{\beta_1}{\sqrt{w_i}} + \beta_2 \frac{X_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \tag{15.2.9}$$

Como puede verificar con facilidad, el término de error transformado en (15.2.9) es homoscedástico. Por consiguiente, después de estimar (15.2.1), ahora podemos calcular (15.2.9) mediante MCO, lo cual no es otra cosa que los *mínimos cuadrados ponderados* (MCP), donde $w_i$ son las ponderaciones.

En teoría, lo que acabamos de describir es correcto; pero en la práctica, se desconoce la verdadera $E(Y_i \mid X_i)$, y en consecuencia, se desconocen las ponderaciones $w_i$. Para calcularlas, utilizamos el siguiente procedimiento, que consta de dos pasos:[5]

**Paso 1.** Efectúe la regresión (15.2.1) por MCO sin considerar el problema de heteroscedasticidad y obtenga $\hat{Y}_i$ = el valor estimado de la verdadera $E(Y_i \mid X_i)$. Luego obtenga $\hat{w}_i = \hat{Y}_i(1 - \hat{Y}_i)$, el valor estimado de $w_i$.

**Paso 2.** Con el $w_i$ estimado transforme los datos como en (15.2.9) y estime la ecuación transformada mediante MCO (es decir, mínimos cuadrados ponderados).

Aunque en breve ilustraremos este procedimiento en nuestro ejemplo, cabe señalar que se pueden usar los errores estándar corregidos por heteroscedasticidad de White para resolver la heteroscedasticidad, siempre que la muestra sea razonablemente grande.

Pero aunque corrijamos la heteroscedasticidad, es necesario resolver primero otro problema que afecta al MLP.

## No cumplimiento de $0 \leq E(Y_i \mid X_i) \leq 1$

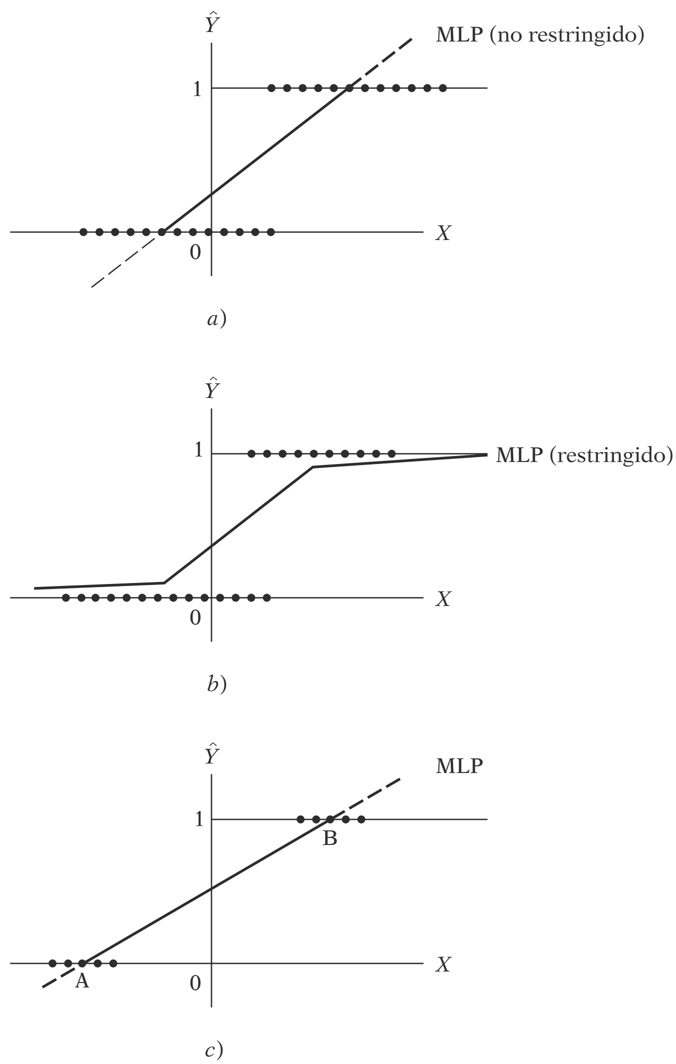
Como $E(Y_i \mid X_i)$ en los modelos lineales de probabilidad mide la probabilidad condicional de que ocurra el suceso $Y$ dado $X$, ésta debe encontrarse necesariamente entre 0 y 1. Aunque *a priori* esto es verdadero, no hay garantía de que $\hat{Y}_i$, los estimadores de $E(Y_i \mid X_i)$, cumplan necesariamente esta restricción, *y éste es el verdadero problema con la estimación del MLP por MCO*. Esto sucede porque MCO no toma en cuenta la restricción $0 \leq E(Y_i) \leq 1$ (una restricción de desigualdad). Hay dos formas de establecer si el $\hat{Y}_i$ estimado se encuentra entre 0 y 1. Una es estimar el MLP mediante el método usual de MCO y determinar si el $\hat{Y}_i$ estimado se encuentra entre 0 y 1. Si algunos valores son menores que 0 (es decir, negativos), para esos casos se supone que $\hat{Y}_i$ es cero; si son mayores que 1, se supone que son 1. El segundo procedimiento es diseñar una técnica de estimación que garantice que las probabilidades condicionales estimadas $\hat{Y}_i$ se encuentren entre 0 y 1. Los modelos logit y probit analizados más adelante garantizarán que las probabilidades estimadas se encuentren con seguridad entre los límites lógicos 0 y 1.

[5] Para conocer la justificación de este procedimiento, véase Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, Nueva York, 1964, pp. 249-250. La justificación consiste básicamente en la de muestras grandes que analizamos en el tema de los mínimos cuadrados generalizados estimados o factibles, en el capítulo sobre heteroscedasticidad (sección 11.6).

## Valor cuestionable de $R^2$ como medida de la bondad del ajuste

$R^2$ calculada de manera convencional tiene un valor limitado en los modelos de respuesta dicótoma. Para ver la razón, considere la figura 15.1. Dado un $X$, $Y$ es igual a 0 o a 1. Por consiguiente, todos los valores de $Y$ se encontrarán en el eje $X$ o en la línea correspondiente a 1. Entonces, por lo general, no se espera que haya un MLP que ajuste bien a tal dispersión, bien sea el MLP *no restringido* (figura 15.1*a*) o el MLP *truncado* o *restringido* (figura 15.1*b*), un MLP estimado en forma tal que no caiga por fuera de la banda lógica 0-1. Como resultado, es probable que el cálculo convencional de $R^2$ sea muy inferior a 1 en estos modelos. En la mayoría de las aplicaciones prácticas, $R^2$ se encuentra entre 0.2 y 0.6. El valor de $R^2$ en ese tipo de modelos será elevado, por ejemplo, superior a 0.8, sólo cuando la dispersión observada esté muy concentrada alrededor de los puntos A y B (figura 15.1*c*), pues en ese caso es fácil modificar la línea uniendo los puntos A y B. En este caso, el valor pronosticado de $Y_i$ estará muy cerca de 0 o de 1.

**FIGURA 15.1**
Modelos lineales de probabilidad.



Por estas razones, John Aldrich y Forrest Nelson sostienen que "debe evitarse el coeficiente de determinación como estadístico de resumen en modelos con variable dependiente cualitativa".[6]

**EJEMPLO 15.1**
*MLP: un ejemplo numérico*

Para ilustrar algunos puntos señalados sobre el MLP en la sección anterior, presentamos un ejemplo numérico. La tabla 15.1 muestra datos inventados sobre propiedad de vivienda $Y$ (1 = tiene casa propia, 0 = no tiene casa propia) e ingreso familiar $X$ (miles de dólares) de 40 familias. Con base en esta información, el MLP estimado por MCO fue el siguiente:

$$\begin{aligned} \hat{Y}_i &= -0.9457 + 0.1021X_i \\ &\quad\;\; (0.1228) \quad\;\; (0.0082) \\ t &= (-7.6984) \quad (12.515) \qquad R^2 = 0.8048 \end{aligned} \tag{15.2.10}$$

Primero, interpretemos esta regresión. El intercepto de −0.9457 da la "probabilidad" de que una familia con ingreso cero tenga una casa propia. Como este valor es negativo y la probabilidad no puede ser negativa, consideramos que este valor es cero, lo cual es razonable en este caso.[7] El valor de la pendiente de 0.1021 significa que para un cambio unitario en el ingreso (aquí, $1 000), en promedio, la probabilidad de tener casa propia aumenta en 0.1021 o alrededor de 10%. Desde luego, con un nivel de ingreso determinado, podemos estimar la probabilidad real de tener casa propia a partir de (15.2.10). Así, para $X = 12$ ($12 000), la probabilidad estimada de tener casa propia es

$$\begin{aligned} (\hat{Y}_i \mid X = 12) &= -0.9457 + 12(0.1021) \\ &= 0.2795 \end{aligned}$$

**TABLA 15.1**
**Datos hipotéticos sobre propiedad de vivienda ($Y = 1$ si tiene casa propia, 0 de lo contrario) e ingreso $X$ (miles de dólares)**

| Familia | $Y$ | $X$ | Familia | $Y$ | $X$ |
|---|---|---|---|---|---|
| 1 | 0 | 8 | 21 | 1 | 22 |
| 2 | 1 | 16 | 22 | 1 | 16 |
| 3 | 1 | 18 | 23 | 0 | 12 |
| 4 | 0 | 11 | 24 | 0 | 11 |
| 5 | 0 | 12 | 25 | 1 | 16 |
| 6 | 1 | 19 | 26 | 0 | 11 |
| 7 | 1 | 20 | 27 | 1 | 20 |
| 8 | 0 | 13 | 28 | 1 | 18 |
| 9 | 0 | 9 | 29 | 0 | 11 |
| 10 | 0 | 10 | 30 | 0 | 10 |
| 11 | 1 | 17 | 31 | 1 | 17 |
| 12 | 1 | 18 | 32 | 0 | 13 |
| 13 | 0 | 14 | 33 | 1 | 21 |
| 14 | 1 | 20 | 34 | 1 | 20 |
| 15 | 0 | 6 | 35 | 0 | 11 |
| 16 | 1 | 19 | 36 | 0 | 8 |
| 17 | 1 | 16 | 37 | 1 | 17 |
| 18 | 0 | 10 | 38 | 1 | 16 |
| 19 | 0 | 8 | 39 | 0 | 7 |
| 20 | 1 | 18 | 40 | 1 | 17 |

(*continúa*)

[6] Aldrick y Nelson, *op. cit.*, p. 15. Para otras medidas de bondad de ajuste en modelos con variables regresadas dicótomas, véase T. Amemiya, "Qualitative Response Models", *Journal of Economic Literature*, vol. 19, 1981, pp. 331-354.

[7] Generalmente, se puede interpretar un valor muy negativo como una probabilidad casi nula de poseer una casa propia cuando el ingreso es cero.

**EJEMPLO 15.1** (*continuación*)

Es decir, la probabilidad de que una familia con un ingreso de $12 000 tenga una casa propia es de alrededor de 28%. La tabla 15.2 muestra las probabilidades estimadas, $\hat{Y}_i$, para los diversos niveles de ingreso enumerados en la tabla. La característica más sobresaliente de esta tabla es que seis valores estimados son negativos y seis valores exceden de uno, lo cual demuestra claramente el punto ya planteado de que, aunque $E(Y_i \mid X_i)$ es positivo y menor que 1, no necesariamente se cumple que sus estimadores $\hat{Y}_i$ sean positivos o inferiores a 1. Ésta es una razón por la cual el MLP no es el modelo recomendado cuando la variable dependiente es dicótoma.

Aunque todos los $Y_i$ estimados fueran positivos e inferiores a 1, el MLP todavía sufre del problema de heteroscedasticidad, lo cual se ve fácilmente de (15.2.8). Como consecuencia, no podemos confiar en los errores estándar estimados que se reportan en (15.12.10). (¿Por qué?) Pero podemos utilizar el procedimiento de mínimos cuadrados ponderados (MCP), ya analizado, para obtener estimaciones más eficientes de los errores estándar. Las ponderaciones necesarias, $\hat{w}_i$, requeridas para la aplicación de MCP se muestran también en la tabla 15.2. Pero observe que algunos $Y_i$ son negativos y otros exceden el valor de uno, los $\hat{w}_i$ correspondientes a estos valores serán negativos. Por tanto, no podemos utilizar estas observaciones en MCP (¿por qué?), con lo cual se reduce el número de observaciones, de 40 a 28 en este ejemplo.[8] Al omitir estas observaciones, la regresión por MCP es

$$\frac{\hat{Y}_i}{\sqrt{\hat{w}_i}} = -1.2456\frac{1}{\sqrt{\hat{w}_i}} + 0.1196\frac{X_i}{\sqrt{\hat{w}_i}} \qquad \textbf{(15.2.11)}$$

$$\begin{aligned} &(0.1206) \qquad (0.0069) \\ t = &(-10.332) \qquad (17.454) \qquad R^2 = 0.9214 \end{aligned}$$

**TABLA 15.2** ***Y* real, *Y* estimado y ponderaciones $w_i$ para el ejemplo de propiedad de vivienda**

| $Y_i$ | $\hat{Y}_i$ | $\hat{w}_i$‡ | $\sqrt{\hat{w}_i}$ | $Y_i$ | $\hat{Y}_i$ | $\hat{w}_i$‡ | $\sqrt{\hat{w}_i}$ |
|---|---|---|---|---|---|---|---|
| 0 | −0.129* | | | 1 | 1.301† | | |
| 1 | 0.688 | 0.2146 | 0.4633 | 1 | 0.688 | 0.2147 | 0.4633 |
| 1 | 0.893 | 0.0956 | 0.3091 | 0 | 0.280 | 0.2016 | 0.4990 |
| 0 | 0.178 | 0.1463 | 0.3825 | 0 | 0.178 | 0.1463 | 0.3825 |
| 0 | 0.280 | 0.2016 | 0.4490 | 1 | 0.688 | 0.2147 | 0.4633 |
| 1 | 0.995 | 0.00498 | 0.0705 | 0 | 0.178 | 0.1463 | 0.3825 |
| 1 | 1.098† | | | 1 | 1.097† | | |
| 0 | 0.382 | 0.2361 | 0.4859 | 1 | 0.893 | 0.0956 | 0.3091 |
| 0 | −0.0265* | | | 0 | 0.178 | 0.1463 | 0.3825 |
| 0 | 0.076 | 0.0702 | 0.2650 | 0 | 0.076 | 0.0702 | 0.2650 |
| 1 | 0.791 | 0.1653 | 0.4066 | 1 | 0.791 | 0.1653 | 0.4055 |
| 1 | 0.893 | 0.0956 | 0.3091 | 0 | 0.382 | 0.2361 | 0.4859 |
| 0 | 0.484 | 0.2497 | 0.4997 | 1 | 1.199† | | |
| 1 | 1.097† | | | 1 | 1.097† | | |
| 0 | −0.333* | | | 0 | 0.178 | 0.1463 | 0.3825 |
| 1 | 0.995 | 0.00498 | 0.0705 | 0 | −0.129* | | |
| 1 | 0.688 | 0.2147 | 0.4633 | 1 | 0.791 | 0.1653 | 0.4066 |
| 0 | 0.076 | 0.0702 | 0.2650 | 1 | 0.688 | 0.2147 | 0.4633 |
| 0 | −0.129* | | | 0 | −0.231* | | |
| 1 | 0.893 | 0.0956 | 0.3091 | 1 | 0.791 | 0.1653 | 0.4066 |

* Se considera igual a cero para evitar que las probabilidades sean negativas.
† Se considera igual a uno para evitar que las probabilidades sean mayores que uno.
‡ $\hat{Y}_i(1 - \hat{Y}_i)$.

[8] Para evitar la pérdida de grados de libertad, podemos dejar que $\hat{Y}_i = 0.01$ cuando los $Y_i$ estimados sean negativos y $\hat{Y}_i = 0.99$ cuando superen o igualen a 1. Véase el ejercicio 15.1.

**EJEMPLO 15.1**
(*continuación*)

Estos resultados demuestran que, en comparación con (15.2.10), los errores estándar estimados son menores y, correspondientemente, las razones $t$ estimadas (en valores absolutos) son más grandes. Pero se debe tomar este resultado con cierta reserva, pues al estimar (15.2.11) se tuvieron que eliminar 12 observaciones. Además, como los $w_i$ son estimaciones, los procedimientos usuales de pruebas de hipótesis estadísticas son válidos, en estricto sentido, en muestras grandes (véase el capítulo 11).

## 15.3 Aplicaciones del MLP

Hasta la aparición de paquetes de computación para estimar los modelos logit y probit (que analizaremos en breve), el MLP era muy común debido a su simplicidad. A continuación ilustramos algunas de estas aplicaciones.

**EJEMPLO 15.2**
*Estudio de Cohen, Rea y Lerman*[9]

En un estudio preparado por el Departamento del Trabajo de Estados Unidos, Cohen, Rea y Lerman examinaron la participación en la fuerza laboral de varias categorías de trabajo como función de diversas variables socioeconómicas y demográficas. En todas sus regresiones, la variable dependiente era dicótoma, con un valor de 1 si la persona pertenecía a la fuerza laboral y de 0 si no pertenecía. En la tabla 15.3 reproducimos una de sus diversas regresiones de variable dependiente dicótoma.

Antes de interpretar los resultados, observe estas características: la regresión anterior se estimó con MCO. Para corregir por heteroscedasticidad, los autores utilizaron el procedimiento de dos etapas, ya descrito, en algunas de sus regresiones, pero encontraron que los errores estándar de las estimaciones así obtenidas no diferían materialmente de las obtenidas sin la corrección por heteroscedasticidad. Tal vez este resultado se deba al tamaño total de la muestra, de alrededor de 25 000. Debido a este gran tamaño de muestra, los valores $t$ estimados pueden probarse por su significancia estadística mediante el procedimiento usual de MCO aunque el término de error adquiera valores dicótomos. El $R^2$ estimado de 0.175 puede parecer relativamente bajo, pero en vista del gran tamaño de la muestra, este $R^2$ aún es significativo con base en la prueba $F$ (véase la sección 8.4). Por último, observe la forma como los autores mezclaron variables cuantitativas y cualitativas, y cómo tomaron en cuenta los efectos de la interacción.

De regreso a la interpretación de los resultados, observamos que cada coeficiente de pendiente da la tasa de cambio en la probabilidad condicional del suceso que ocurre ante un cambio unitario en el valor de la variable explicativa. Por ejemplo, el coeficiente de −0.2753 que acompaña a la variable "más de 65 años" significa que, si se mantienen constantes todos los demás factores, la probabilidad de participación en la fuerza laboral de mujeres en este grupo de edad es menor en alrededor de 27% (en comparación con la categoría base de mujeres con edades entre 22 y 54 años). Con el mismo razonamiento, el coeficiente de 0.3061, asociado a la variable "más de 16 años de escolaridad", significa que, si se mantienen constantes todos los demás factores, la probabilidad de que las mujeres con esta misma escolaridad participen en la fuerza laboral es más alta en cerca de 31% (comparado con la categoría base, que son mujeres con menos de cinco años de escolaridad).

Ahora, considere el **término de interacción** estado civil y edad. La tabla muestra que la probabilidad de participación en la fuerza laboral es más alta en cerca de 29% para las mujeres solteras (comparado con la categoría base) y más baja en alrededor de 28% para las mujeres de más de 65 años de edad (de nuevo, en relación con la categoría base). Pero la probabilidad de participación de mujeres solteras y mayores de 65 años es menor en cerca de 20% comparada con la categoría base. Esto implica que es probable que las mujeres solteras mayores de 65 años participen en la fuerza laboral en mayor proporción que las casadas o clasificadas en la categoría "otros" que tienen más de 65 años de edad.

(*continúa*)

[9] Malcolm S. Cohen, Samuel A. Rea, Jr., y Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, Departamento de Trabajo de Estados Unidos, 1970.

**EJEMPLO 15.2**
(*continuación*)

**TABLA 15.3 Participación en la fuerza laboral**
**Regresión de mujeres, mayores de 22 años, que viven en las 96 áreas estadísticas metropolitanas estándar (AEME) más grandes (variable dependiente: perteneció o no a la fuerza laboral durante 1966)**

| Variable explicativa | Coeficiente | Razón *t* |
|---|---|---|
| Constante | 0.4368 | 15.4 |
| Estado civil | | |
| Casada, vive con el esposo | — | — |
| Casada, otro | 0.1523 | 13.8 |
| Soltera | 0.2915 | 22.0 |
| Edad | | |
| 22-54 | — | — |
| 55-64 | –0.0594 | –5.7 |
| más de 65 | –0.2753 | –9.0 |
| Años de escolaridad | | |
| 0-4 | — | — |
| 5-8 | 0.1255 | 5.8 |
| 9-11 | 0.1704 | 7.9 |
| 12-15 | 0.2231 | 10.6 |
| Más de 16 | 0.3061 | 13.3 |
| Tasa de desempleo (1966), % | | |
| Menos de 2.5 | — | — |
| 2.5-3.4 | –0.0213 | –1.6 |
| 3.5-4.0 | –0.0269 | –2.0 |
| 4.1-5.0 | –0.0291 | –2.2 |
| Más de 5.1 | –0.0311 | –2.4 |
| Cambio en el empleo (1965-1966), % | | |
| Menor que 3.5 | — | — |
| 3.5-6.49 | 0.0301 | 3.2 |
| Más de 6.5 | 0.0529 | 5.1 |
| Oportunidades relativas de empleo, % | | |
| Menos de 62 | — | — |
| 62-73.9 | 0.0381 | 3.2 |
| Más de 74 | 0.0571 | 3.2 |
| IFMJ, $ | | |
| Menos de 1 500 y negativo | — | — |
| 1 500-7 499 | –0.1451 | –15.4 |
| 7 500 y más | –0.2455 | –24.4 |
| Interacción (estado civil y edad) | | |
| Estado civil — Edad | | |
| Otro — 55-64 | –0.0406 | –2.1 |
| Otro — Más de 65 | –0.1391 | –7.4 |
| Soltera — 55-64 | –0.1104 | –3.3 |
| Soltera — Más de 65 | –0.2045 | –6.4 |
| Interacción (edad y años de escolaridad terminados) | | |
| Edad — Años de escolaridad | | |
| Más de 65 — 5-8 | –0.0885 | –2.8 |
| Más de 65 — 9-11 | –0.0848 | –2.4 |
| Más de 65 — 12-15 | –0.1288 | –4.0 |
| Más de 65 — 16 y más | –0.1628 | –3.6 |
| $R^2 = 0.175$ | | |
| Núm. de observaciones = 25 153 | | |

*Nota:* — indica la categoría base o categoría omitida.
IFMJ: Ingreso familiar menos jornales propios e ingresos por salarios.

Fuente: Malcolm S. Cohen, Samuel A. Rea, Jr., y Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, Departamento de trabajo de Estados Unidos, 1970, tabla F-6, pp. 212-213.

**EJEMPLO 15.2** (*continuación*)

Con este procedimiento puede interpretar sin dificultad el resto de los coeficientes dados en la tabla 15.3. Con esta información es fácil obtener las estimaciones de las probabilidades condicionales de la participación de la fuerza laboral de las diversas categorías. Así, si deseamos encontrar la probabilidad para mujeres casadas (otras), entre 22 y 54 años de edad, con 12 a 15 años de escolaridad, con una tasa de desempleo de 2.4 a 3.4%, cambio de empleo de 3.5 a 6.49%, oportunidades relativas de empleo de 74% y por encima y con un IFMJ de $7 500 y más, obtenemos

$$0.4368 + 0.1523 + 0.2231 - 0.0213 + 0.0301 + 0.0571 - 0.2455 = 0.6326$$

En otras palabras, la probabilidad de la participación de las mujeres en la fuerza laboral con las características anteriores se estima en alrededor de 63%.

**EJEMPLO 15.3** *Predicción de la clasificación de bonos*

Con base en datos de series de tiempo agrupadas y de corte transversal de 200 bonos Aa (alta calidad) y Baa (calidad media) de 1961 a 1966, Joseph Cappelleri estimó el siguiente modelo de predicción para la clasificación de bonos.[10]

$$Y_i = \beta_1 + \beta_2 X_{2i}^2 + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

donde $Y_i$ = 1 si la clasificación del bono es Aa (clasificación de la empresa Moody)
= 0 si la clasificación del bono es Baa (clasificación de la empresa Moody)
$X_2$ = razón de capitalización de la deuda, una medida de apalancamiento

$$= \frac{\text{valor en dólares de la deuda de largo plazo}}{\text{valor en dólares de la capitalización total}} \cdot 100$$

$X_3$ = tasa de rentabilidad

$$= \frac{\text{valor en dólares del ingreso después de impuestos}}{\text{valor en dólares de los activos totales netos}} \cdot 100$$

$X_4$ = desviación estándar de la tasa de rentabilidad, una medida de la variabilidad de la tasa de rentabilidad
$X_5$ = activos totales netos (miles de dólares), una medida del tamaño

*A priori*, se espera que $\beta_2$ y $\beta_4$ sean negativos (¿por qué?) y que $\beta_3$ y $\beta_5$ positivos.

Después de corregir por heteroscedasticidad y por autocorrelación de primer orden, Cappelleri obtuvo los siguientes resultados:[11]

$$\begin{aligned} \hat{Y}_i = {} & \underset{(0.1775)}{0.6860} - \underset{(0.0024)}{0.0179} X_{2i}^2 + \underset{(0.0486)}{0.0486} X_{3i} + \underset{(0.0178)}{0.0572} X_{4i} + \underset{(0.039)(E\text{-}8)}{0.378(E\text{-}7)} X_5 \\ & R^2 = 0.6933 \end{aligned} \qquad \textbf{(15.3.1)}$$

*Nota:* 0.378 (*E*-7) significa 0.0000000378, etcétera.

Todos los coeficientes, a excepción del asociado a $X_4$, tienen los signos correctos. Dejamos a los estudiantes de finanzas que deduzcan la razón por la cual el coeficiente de la variabilidad de la tasa de rentabilidad tiene signo positivo, pues se espera que cuanto mayor sea la variabilidad en las utilidades, menos probable es que la empresa Moody dé una clasificación Aa, si las demás condiciones se mantienen iguales.

La interpretación de la regresión es clara. Por ejemplo, el 0.0486 asociado a $X_3$ significa que, si las demás condiciones se mantienen iguales, un incremento de un punto porcentual en la tasa de rendimiento generará, en promedio, alrededor de un 0.05 de incremento en la probabilidad de que un bono obtenga la clasificación Aa. En forma similar, cuanto más alta sea la tasa de endeudamiento elevada al cuadrado, menor será la probabilidad (en 0.02) de que un bono reciba la clasificación Aa por unidad de incremento en esta tasa.

[10] Joseph Cappelleri, "Predicting a Bond Rating", documento universitario sin publicar, C.U.N.Y. El modelo empleado es una modificación del que aparece en Thomas F. Pogue y Robert M. Soldofsky, "What Is in a Bond Rating?", *Journal of Financial and Quantitative Analysis*, junio de 1969, pp. 201-228.

[11] Algunas probabilidades estimadas antes de corregir por heteroscedasticidad fueron negativas y otras fueron superiores a 1; para facilitar el cálculo de las ponderaciones $w_i$ en estos casos, se supuso que fueron 0.01 y 0.99, respectivamente.

**EJEMPLO 15.4**
*¿Quién tiene una tarjeta de débito?*

Como ocurre con las tarjetas de crédito, en la actualidad es muy común que los consumidores usen las de débito. Los comerciantes las prefieren porque cuando se usa una tarjeta de débito, la cantidad que uno compra se deduce automáticamente de la cuenta de cheques u otra cuenta designada por el usuario. Para averiguar qué factores determinan el uso de la tarjeta de débito, obtuvimos datos sobre 60 clientes y consideramos el siguiente modelo:[12]

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

donde $Y = 1$ para quien posee una tarjeta de débito, 0 en otro caso; $X_2 =$ saldo de la cuenta en dólares; $X_3 =$ número de transacciones en cajeros automáticos; $X_4 = 1$ si la cuenta devenga intereses, 0 en caso contrario.

En vista de que el modelo lineal de probabilidad (MLP) muestra heteroscedasticidad, presentamos los resultados habituales de MCO y los resultados de MCO corregidos por heteroscedasticidad en forma tabular.

| **Variable** | **Coeficiente** | **Coeficiente*** |
|---|---|---|
| Constante | 0.3631 | 0.3631 |
| | (0.1796)** | (0.1604)** |
| Saldo | 0.00028** | 0.00028** |
| | (0.00015) | (0.00014) |
| Cajero automático | –0.0269 | –0.0269 |
| | (0.208) | (0.0202) |
| Interés | –0.3019** | –0.3019** |
| | (0.1448) | (0.1353) |
| $R^2$ | 0.1056 | (0.1056) |

*Nota:* * denota errores estándar corregidos por heteroscedasticidad.
** significativo en el nivel de 5% aproximadamente.

Como indican estos resultados, los usuarios con saldos más altos en sus cuentas tienden a tener tarjeta de débito. Cuanto más alta es la tasa de interés que se paga sobre el saldo de la cuenta, menor es la tendencia a tener tarjeta de débito. Aunque la variable cajero automático no es significativa, observe que tiene signo negativo. Esto se debe quizá a las comisiones por transacción en cajero automático.

No existe una gran diferencia entre los errores estándar estimados con y sin corrección por heteroscedasticidad. Para ahorrar espacio, no presentamos los valores ajustados (es decir, las probabilidades estimadas), pero todas se sitúan entre los límites de 0 y 1. Sin embargo, no hay garantía de que así ocurra en todos los casos.

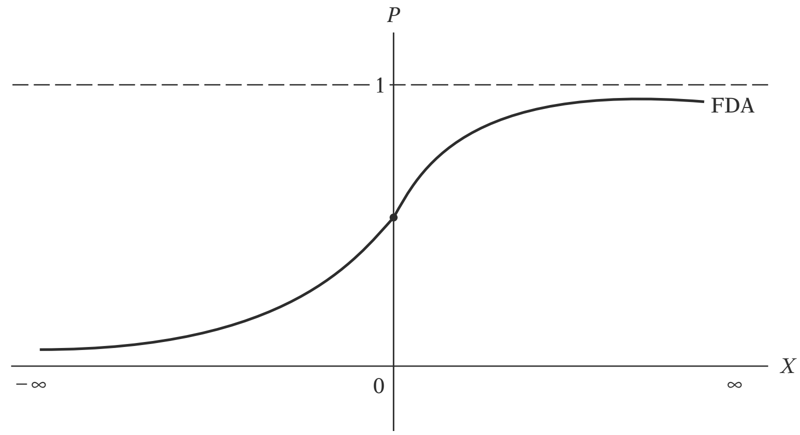
## 15.4 Alternativas al MLP

Como vimos, el MLP tiene infinidad de problemas, como 1) la no normalidad de los $u_i$, 2) la heteroscedasticidad de $u_i$, 3) la posibilidad de que $\hat{Y}_i$ se encuentre fuera del rango 0-1 y 4) los valores generalmente bajos de $R^2$. Pero estos problemas son superables. Por ejemplo, se pueden utilizar MCP para resolver el problema de heteroscedasticidad o incrementar el tamaño de la muestra y reducir así el problema de la no normalidad. Mediante técnicas de mínimos cuadrados restringidos o de programación matemática, es posible hacer que las probabilidades estimadas se encuentren dentro del intervalo 0-1.

Pero incluso así, el problema fundamental con el MLP es que lógicamente no es un modelo muy atractivo porque supone que $P_i = E(Y = 1 \mid X)$ aumenta linealmente con $X$, es decir, el efecto marginal o incremental de $X$ permanece constante todo el tiempo. Así, en el ejemplo de propiedad de vivienda encontramos que, a medida que $X$ aumenta en una unidad ($ 1 000), la probabili-

[12] Los datos empleados en el análisis se obtuvieron de Douglas A. Lind, William G. Marchal y Robert D. Mason, *Statistical Techniques in Business and Economics,* 11a. ed., McGraw-Hill, 2002, apéndice *N,* pp. 775-776. No utilizamos todas las variables que emplearon los autores.

**FIGURA 15.2**
Función de distribución acumulativa (FDA).




dad de ser propietario de una casa aumenta en la misma cantidad constante de 0.10. Esto es así independientemente de que el nivel del ingreso sea de \$8 000, \$10 000, \$18 000 o \$22 000. Esto no sucede jamás en la realidad. En verdad se esperaría que $P_i$ estuviera relacionado en forma no lineal con $X_i$: con ingresos muy bajos, una familia no será propietaria de una casa, pero en un nivel de ingresos lo bastante altos, por ejemplo, $X^*$, es muy probable que sí tenga casa propia. Cualquier incremento en el ingreso más allá de $X^*$ tendrá un efecto pequeño sobre la probabilidad de tener casa propia. Así, en ambos extremos de la distribución de ingresos, la probabilidad de ser dueño de una casa prácticamente no se verá afectada por un pequeño incremento en $X$.

Por consiguiente, lo que necesitamos es un modelo (probabilístico) que tenga estas dos características: 1) a medida que aumente $X_i$, $P_i = E(Y = 1 \mid X)$ también aumente pero nunca se salga del intervalo 0-1, y 2) la relación entre $P_i$ y $X_i$ sea no lineal, es decir, "uno se acerca a cero con tasas cada vez más lentas a medida que se reduce $X_i$, y se acerca a uno con tasas cada vez más lentas a medida que $X_i$ se hace muy grande".[13]

En términos geométricos, el modelo que deseamos tendría la forma de la figura 15.2. Observe en este modelo que la probabilidad se encuentra entre 0 y 1, y que éste varía en forma no lineal con $X$.

El lector se dará cuenta de que la curva en forma de S, o sigmoidea, en la figura se parece mucho a la **función de distribución acumulativa** de una variable aleatoria (FDA).[14] Por consiguiente, se puede utilizar fácilmente la FDA en regresiones de modelos en los cuales la variable de respuesta es dicótoma, para adquirir valores 0-1. La pregunta práctica ahora es, ¿cuál FDA?: aunque todas las FDA tienen forma de S, para cada variable aleatoria hay una FDA única. Por razones tanto históricas como prácticas, las FDA que suelen seleccionarse para representar los modelos de respuesta 0-1 son 1) la logística y 2) la normal; la primera da lugar al modelo **logit**, y la última, al modelo **probit** (o **normit**).

Aunque el análisis detallado de los modelos logit y probit trasciende el alcance de esta obra, indicaremos de manera un poco informal el modo de estimar e interpretar tales modelos.

## 15.5 El modelo logit

Continuamos con el ejemplo de propiedad de vivienda para explicar las ideas básicas del modelo logit. Recuerde que en la explicación de la propiedad de vivienda en relación con el ingreso, el MLP fue

$$P_i = \beta_1 + \beta_2 X_i \tag{15.5.1}$$

[13] John Aldrich y Forrest Nelson, *op. cit.*, p. 26.

[14] Como analizamos en el **apéndice A**, la FDA de una variable aleatoria $X$ es sencillamente la probabilidad de que adopte un valor menor o igual a $x_0$, donde $x_0$ es algún valor numérico especificado de $X$. En resumen, $F(X)$, la FDA de $X$, es $F(X = x_0) = P(X \leq x_0)$.

donde $X$ es el ingreso y $P_i = E(Y_i = 1 \mid X_i)$ significa que la familia es propietaria de una casa. Pero considere ahora la siguiente representación de la propiedad de vivienda:

$$P_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \tag{15.5.2}$$

Para facilidad de la exposición, escribimos (15.5.2) como

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^Z}{1 + e^Z} \tag{15.5.3}$$

donde $Z_i = \beta_1 + \beta_2 X_i$.

La ecuación (15.5.3) representa lo que se conoce como **función de distribución logística** (acumulativa).[15]

Es fácil verificar que a medida que $Z_i$ se encuentra dentro de un rango de $-\infty$ a $+\infty$, $P_i$ se encuentra dentro de un rango de 0 a 1, y que $P_i$ no está linealmente relacionado con $Z_i$ (es decir, con $X_i$), lo que satisface los dos requisitos considerados antes.[16] Pero parece que al satisfacer estos requisitos creamos un problema de estimación, porque $P_i$ es no lineal no sólo en $X$ sino también en las $\beta$, como se ve a partir de (15.5.2). Esto significa que no podemos estimar los parámetros con el procedimiento habitual de MCO.[17] Pero este problema es más aparente que real porque (15.5.2) puede linealizarse, lo cual se demuestra de la siguiente manera.

Si $P_i$, la probabilidad de tener casa propia, está dada por (15.5.3), entonces $(1 - P_i)$, la probabilidad de no tener casa propia, es

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \tag{15.5.4}$$

Por consiguiente, podemos escribir

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \tag{15.5.5}$$

Ahora $P_i/(1 - P_i)$ es sencillamente la **razón de las probabilidades** en favor de tener una casa propia: la razón de la probabilidad de que una familia posea una casa propia respecto de la probabilidad de que no la posea. Así, si $P_i = 0.8$, significa que las posibilidades son 4 a 1 en favor de que la familia posea una casa propia.

Ahora, si tomamos el logaritmo natural de (15.5.5), obtenemos un resultado muy interesante, a saber,

$$\begin{aligned} L_i = \ln\left(\frac{P_i}{1 - P_i}\right) &= Z_i \\ &= \beta_1 + \beta_2 X_i \end{aligned} \tag{15.5.6}$$

[15] El modelo logístico es muy común en análisis de fenómenos de crecimiento, como población, PNB, oferta monetaria, etc. Para conocer los detalles teóricos y prácticos de los modelos logit y probit, véase J.S. Kramer, *The Logit Model for Economists*, Edward Arnold, Londres, 1991; y G.S. Maddala, *op. cit.*

[16] Observe que a medida que $Z_i \to +\infty$, $e^{-Z_i}$ tiende a cero, y a medida que $Z_i \to -\infty$, $e^{-Z_i}$ aumenta indefinidamente. Recuerde que $e = 2.71828$.

[17] Desde luego, se pueden utilizar las técnicas de estimación no lineales analizadas en el capítulo 14. Véase también la sección 15.8.

es decir, $L$, el logaritmo de la razón de las probabilidades, no es sólo lineal en $X$, sino también (desde el punto de vista de estimación) lineal en los parámetros.[18] $L$ se llama **logit**, y de aquí el nombre **modelo logit** para modelos como (15.5.6).

Observe estas características del modelo logit.

1. A medida que $P$ va de 0 a 1 (es decir, a medida que $Z$ varía de $-\infty$ a $+\infty$, el logit $L$ va de $-\infty$ a $+\infty$. Es decir, aunque las probabilidades (por necesidad) se encuentran entre 0 y 1, los logit no están acotados en esa forma.

2. Aunque $L$ es lineal en $X$, las probabilidades en sí mismas no lo son. Esta propiedad contrasta con el MLP (15.5.1), en donde las probabilidades aumentan linealmente con $X$.[19]

3. Aunque en el modelo anterior incluimos sólo una variable $X$, o regresora, podemos añadir tantas regresoras como indique la teoría subyacente.

4. Si $L$, el logit, es positivo, significa que cuando se incrementa el valor de la(s) regresora(s), aumentan las posibilidades de que la regresada sea igual a 1 (lo cual indica que sucederá algo de interés). Si $L$ es negativo, las posibilidades de que la regresada iguale a 1 disminuyen conforme se incrementa el valor de $X$. Para expresarlo de otra forma, el logit se convierte en negativo y se incrementa en gran medida conforme la razón de las probabilidades disminuye de 1 a 0; además, se incrementa en gran medida y se vuelve positivo conforme la razón de las probabilidades aumenta de 1 a infinito.[20]

5. De manera más formal, la interpretación del modelo dado en (15.5.6) es la siguiente: $\beta_2$, la pendiente, mide el cambio en $L$ ocasionado por un cambio unitario en $X$, es decir, dice cómo cambia el logaritmo de las posibilidades en favor de tener una casa propia a medida que el ingreso cambia en una unidad, por ejemplo, \$1 000. El intercepto $\beta_1$ es el valor del logaritmo de las posibilidades en favor de tener una casa propia si el ingreso es cero. Al igual que la mayoría de las interpretaciones de los interceptos, esta interpretación puede no tener significado físico alguno.

6. Con un nivel determinado de ingreso, por ejemplo, $X^*$, si de verdad deseamos estimar la probabilidad misma de tener una casa propia, y no las posibilidades en favor de tener una casa propia, podemos hacerlo directamente a partir de (15.5.3) una vez que dispongamos de las estimaciones de $\beta_1$ y $\beta_2$. Pero esto plantea la pregunta más importante: ¿cómo estimar $\beta_1$ y $\beta_2$ en primer lugar? La respuesta está en la siguiente sección.

7. Mientras que el MLP supone que $P_i$ está linealmente relacionado con $X_i$, el modelo logit supone que el logaritmo de la razón de probabilidades está relacionado linealmente con $X_i$.

## 15.6 Estimación del modelo logit

Para fines de estimación, escribimos (15.5.6) de la siguiente manera:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 X_i + u_i \qquad \textbf{(15.6.1)}$$

En breve analizaremos las propiedades del término de error estocástico $u_i$.

[18] Recuerde que el supuesto de linealidad de MCO no requiere que la variable $X$ sea necesariamente lineal. Así, podemos tener $X^2$, $X^3$, etc., como regresoras en el modelo. Para nuestro propósito, lo crucial es la linealidad en los parámetros.

[19] Mediante cálculo, se demuestra que $dP/dX = \beta_2 P(1 - P)$, lo cual muestra que la tasa de cambio de la probabilidad respecto de $X$ contiene no sólo a $\beta_2$, sino también al nivel de probabilidad a partir del cual se mide el cambio (véase más al respecto en la sección 15.7). A propósito, observe que el efecto de un cambio unitario en $X_i$ sobre $P$ es máximo cuando $P = 0.5$ y mínimo cuando $P$ está cercano a 0 o a 1.

[20] Esta observación la hizo David Garson.

Para estimar (15.6.1), además de $X_i$, necesitamos los valores de la regresada, o del logit, $L_i$. Esto depende del tipo de datos que se analicen. Éstos se clasifican en dos categorías: 1) *datos de nivel individual, o micro*, y 2) *datos agrupados o duplicados*.

## Datos de nivel individual

Si disponemos de datos sobre familias individuales, como en la tabla 15.1, no es factible la estimación de (15.6.1) por MCO, lo cual es fácil de ver. En términos de los datos proporcionados en la tabla 15.1, $P_i = 1$ si una familia es dueña de una casa y $P_i = 0$ si no tiene casa propia. Pero si colocamos estos valores directamente en el logit $L_i$, obtenemos

$$L_i = \ln\left(\frac{1}{0}\right) \qquad \text{si una familia es dueña de una casa}$$

$$L_i = \ln\left(\frac{0}{1}\right) \qquad \text{si una familia no es dueña de una casa}$$

Obvio, estas expresiones no tienen sentido. Por consiguiente, si la información disponible está en un nivel micro o individual, no podemos estimar (15.6.1) mediante la rutina de MCO estándar. En esta situación quizá debamos recurrir al método de **máxima verosimilitud (MV)** para estimar los parámetros. Aunque analizamos los rudimentos de este método en el apéndice del capítulo 4, su aplicación, en el presente contexto, se presenta en el apéndice 15A, sección 15A.1, como consideración a los lectores que deseen aprender más sobre él.[21] Programas como MICROFIT, *EViews*, LIMDEP, SHAZAM, PC-GIVE, STATA y MINITAB tienen rutinas internas para estimar el modelo logit en el nivel individual. Ilustraremos el método de VM más adelante en este capítulo.

## Datos agrupados o duplicados

Ahora considere los datos de la tabla 15.4. Esta tabla proporciona datos *agrupados* o *duplicados* (observaciones repetidas) sobre diversas familias, de acuerdo con el nivel de ingreso y el número de familias que tienen casa propia en cada nivel de ingreso. Correspondiente a cada nivel de ingreso $X_i$, hay $N_i$ familias, de las cuales $n_i$ tienen casa propia ($n_i \leq N_i$). Por consiguiente, si ahora calculamos

$$\hat{P}_i = \frac{n_i}{N_i} \tag{15.6.2}$$

**TABLA 15.4**
**Datos hipotéticos sobre $X_i$ (ingreso), $N_i$ (número de familias con ingreso $X_i$) y $n_i$ (número de familias que tienen casa propia)**

| X (miles de dólares) | $N_i$ | $n_i$ |
|---|---|---|
| 6 | 40 | 8 |
| 8 | 50 | 12 |
| 10 | 60 | 18 |
| 13 | 80 | 28 |
| 15 | 100 | 45 |
| 20 | 70 | 36 |
| 25 | 65 | 39 |
| 30 | 50 | 33 |
| 35 | 40 | 30 |
| 40 | 25 | 20 |

[21] Para un análisis de máxima verosimilitud comparativamente sencillo en el contexto del modelo logit, véase John Aldrich y Forrest Nelson, *op. cit.*, pp. 49-54. Véase también Alfred Demaris, *Logit Modeling: Practical Applications*, Sage, Newbury Park, California, 1992.

es decir, *frecuencia relativa*, podemos utilizarla como una estimación de la verdadera $P_i$ correspondiente a cada $X_i$. Si $N_i$ es relativamente grande, $\hat{P}_i$ será una estimación razonablemente buena de $P_i$.[22] Con la $P_i$ estimada, podemos obtener el logit estimado como

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{15.6.3}$$

lo cual será una estimación relativamente buena del verdadero logit $L_i$, si el número de observaciones $N_i$ en cada nivel $X_i$ es razonablemente grande.

En resumen, con la información *agrupada* o *replicada* (observaciones repetidas), como la de la tabla 15.4, podemos obtener información sobre la variable dependiente, los logit, para estimar el modelo (15.6.1). ¿Podemos entonces aplicar MCO a (15.6.3) y estimar los parámetros en la forma usual? La respuesta es: aún no, pues hasta el momento no hemos dicho nada sobre las propiedades del término de perturbación estocástico. Puede demostrarse que si $N_i$ es relativamente grande, y si cada observación en una clase de ingreso dado $X_i$ está distribuida en forma independiente como una variable binomial, entonces

$$u_i \sim N\left[0, \frac{1}{N_i P_i(1-P_i)}\right] \tag{15.6.4}$$

es decir, $u_i$ sigue una distribución normal con media cero y varianza igual a $1/[N_i P_i(1-P_i)]$.[23]

Por consiguiente, como en el caso del MLP, el término de perturbación en el modelo logit es heteroscedástico. Así, en lugar de MCO, debemos utilizar mínimos cuadrados ponderados (MCP). Para fines empíricos; sin embargo, reemplazaremos la $P_i$ desconocida por $\hat{P}_i$ y utilizaremos

$$\hat{\sigma}^2 = \frac{1}{N_i \hat{P}_i(1-\hat{P}_i)} \tag{15.6.5}$$

como estimador de $\sigma^2$.

Ahora describiremos los diversos pasos en la estimación de la regresión logit (15.6.1):

1. Para cada nivel de ingreso $X$, calcule la probabilidad estimada de tener una casa propia como $\hat{P}_i = n_i/N_i$.
2. Por cada $X_i$, obtenga el logit mediante[24]

$$\hat{L}_i = \ln[\hat{P}_i/(1-\hat{P}_i)]$$

3. Para resolver el problema de heteroscedasticidad, transforme (15.6.1) de la siguiente manera:[25]

$$\sqrt{w_i} L_i = \beta_1\sqrt{w_i} + \beta_2\sqrt{w_i} X_i + \sqrt{w_i} u_i \tag{15.6.6}$$

[22] De la estadística elemental recuerde que la probabilidad de un evento es el límite de la frecuencia relativa a medida que el tamaño de la muestra se hace infinitamente grande.

[23] Como se demuestra en la teoría de probabilidad elemental, $\hat{P}_i$, la proporción de éxitos (en este caso, la propiedad de una casa), sigue la distribución binomial con media igual a la verdadera $P_i$ y varianza igual a $P_i(1-P_i)/N_i$; y a medida que $N_i$ aumenta indefinidamente, la distribución binomial se aproxima a la distribución normal. Las propiedades distributivas de $u_i$ dadas en (15.6.4) se desprenden de esta teoría básica. Hay mayores detalles en Henry Theil, "On the Relationships Involving Qualitative Variables", *American Journal of Sociology*, vol. 76, julio de 1970, pp. 103-154.

[24] Como $\hat{P}_i = n_i/N_i$, $L_i$ se expresa alternativamente como $\hat{L}_i = \ln n_i/(N_i - n_i)$. A propósito, debe resaltarse que para evitar que $\hat{P}_i$ tome el valor de 0 o de 1, en la práctica $\hat{L}_i$ se mide como $\hat{L}_i = \ln(n_i + \frac{1}{2})/(N_i - n_i + \frac{1}{2}) = \ln(\hat{P}_i + 1/2N_i)/(1 - \hat{P}_i + 1/2N_i)$. Se recomienda, como regla práctica, que $N_i$ sea por lo menos 5 para cada valor de $X_i$. Hay mayores detalles en D.R. Cox, *Analysis of Binary Data*, Methuen, Londres, 1970, p. 33.

[25] Si estimamos (15.6.1) sin tener en cuenta la heteroscedasticidad, los estimadores, aunque sean insesgados, no serán eficientes, como sabemos por el capítulo 11.

que escribimos como

$$L_i^* = \beta_1\sqrt{w_i} + \beta_2 X_i^* + v_i \tag{15.6.7}$$

donde las ponderaciones $w_i = N_i\hat{P}_i(1-\hat{P}_i)$; $L_i^* =$ transformada o ponderada $L_i$; $X_i^* = X_i$ transformada o ponderada, y $v_i =$ término de error transformado. Es fácil verificar que el término de error transformado $v_i$ es homoscedástico, sin olvidar que la varianza de error original es $\sigma_u^2 = 1/[N_i P_i(1-P_i)]$.

4. Estime (15.6.6) mediante MCO; recuerde que MCP es MCO aplicado sobre los datos transformados. Observe que en (15.6.6) no hay término de intercepto introducido explícitamente (¿por qué?). Por consiguiente, se tendrá que estimar (15.6.6) mediante el procedimiento de regresión a través del origen.
5. Establezca intervalos de confianza y/o pruebas de hipótesis dentro del marco usual de MCO, *pero tenga en mente que todas las conclusiones serán válidas en estricto sentido si la muestra es razonablemente grande* (¿por qué?). Por consiguiente, en muestras pequeñas, los resultados estimados deben interpretarse con cautela.

## 15.7 Modelo logit agrupado (glogit): ejemplo numérico

Para ilustrar la teoría que acabamos de analizar, utilizaremos los datos de la tabla 15.4. Como están agrupados, el modelo logit que se basa en tales datos se llamará modelo logit agrupado, o glogit, para abreviar. Los datos necesarios y otros cálculos pertinentes necesarios para instrumentar el modelo glogit se dan en la tabla 15.5. Los resultados de la regresión con mínimos cuadrados ponderados (15.6.7) que se basa en los datos de la tabla 15.5 son los siguientes: observe que no hay intercepto en la ecuación (15.6.7); por consiguiente, en este caso resulta apropiado el procedimiento de regresión a través del origen.

$$\begin{aligned} \hat{L}_i^* &= \quad -1.59474\sqrt{w_i} + \quad 0.07862X_i^* \\ \text{ee} &= \quad (0.11046) \qquad\quad (0.00539) \\ t &= (-14.43619) \qquad (14.56675) \qquad R^2 = 0.9642 \end{aligned} \tag{15.7.1}$$

$R^2$ es el coeficiente de correlación al cuadrado entre la $L_i^*$ verdadera y la estimada. $L_i^*$ y $X_i^*$ son la $L_i$ y la $X_i$ ponderadas, como se muestra en (15.6.6). Aunque presentamos los cálculos del modelo logit agrupado en la tabla 15.5 por razones pedagógicas, esto se efectúa fácilmente si se pulsa el comando **glogit** (logit agrupado) en STATA.

### Interpretación del modelo logit estimado

¿Cómo interpretamos (15.7.1)? Hay diversas formas, algunas intuitivas y otras no, a saber:

*Interpretación Logit*

Como lo muestra (15.7.1), el coeficiente de pendiente estimado indica que para un incremento unitario ($1 000) en el ingreso ponderado, el logaritmo ponderado de las posibilidades en favor de tener casa propia aumenta en alrededor de 0.08. Esta interpretación mecánica no resulta muy atractiva.

*Interpretación de las probabilidades*

Recuerde que $L_i = \ln\,[P_i/(1-P_i)]$. Por consiguiente, al tomar el antilogaritmo del logit estimado, se tiene $P_i/(1-P_i)$, es decir, la razón de las probabilidades. En consecuencia, al tomar el antilog de (15.7.1), obtenemos:

**TABLA 15.5** **Datos para estimar el modelo logit sobre la propiedad de vivienda**

| $X$ (miles de dólares) | $N_i$ | $n_i$ | $\hat{P}_i$ | $1 - \hat{P}_i$ | $\frac{\hat{P}_i}{1 - \hat{P}_i}$ | $\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1 - \hat{P}_i}\right)$ | $N_i \hat{P}_i (1 - \hat{P}_i) = w_i$ | $\sqrt{w_i} = \sqrt{N_i \hat{P}_i (1 - \hat{P}_i)}$ | $\hat{L}_i^* = \hat{L}_i \sqrt{w_i}$ | $\hat{X}_i^* = \hat{X}_i \sqrt{w_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) = (3) ÷ (2) | (5) | (6) | (7) | (8) | (9) = $\sqrt{(8)}$ | (10) = (7)(9) | (11) = (1)(9) |
| 6 | 40 | 8 | 0.20 | 0.80 | 0.25 | −1.3863 | 6.40 | 2.5298 | −3.5071 | 15.1788 |
| 8 | 50 | 12 | 0.24 | 0.76 | 0.32 | −1.1526 | 9.12 | 3.0199 | −3.4807 | 24.1592 |
| 10 | 60 | 18 | 0.30 | 0.70 | 0.43 | −0.8472 | 12.60 | 3.5496 | −3.0072 | 35.4960 |
| 13 | 80 | 28 | 0.35 | 0.65 | 0.54 | −0.6190 | 18.20 | 4.2661 | −2.6407 | 55.4593 |
| 15 | 100 | 45 | 0.45 | 0.55 | 0.82 | −0.2007 | 24.75 | 4.9749 | −0.9985 | 74.6235 |
| 20 | 70 | 36 | 0.51 | 0.49 | 1.04 | 0.0570 | 17.49 | 4.1816 | 0.1673 | 83.6506 |
| 25 | 65 | 39 | 0.60 | 0.40 | 1.50 | 0.4054 | 15.60 | 3.9497 | 1.6012 | 98.7425 |
| 30 | 50 | 33 | 0.66 | 0.34 | 1.94 | 0.6633 | 11.20 | 3.3496 | 2.2218 | 100.4880 |
| 35 | 40 | 30 | 0.75 | 0.25 | 3.0 | 1.0986 | 7.50 | 2.7386 | 3.0086 | 95.8405 |
| 40 | 25 | 20 | 0.80 | 0.20 | 4.0 | 1.3863 | 4.00 | 2.000 | 2.7726 | 80.0000 |

$$\frac{\hat{P}_i}{1-\hat{P}_i} = e^{-1.59474\sqrt{w_i}+0.07862X_i^*} \qquad (15.7.2)$$
$$= e^{-1.59474\sqrt{w_i}} \cdot e^{0.07862X_i^*}$$

Mediante una calculadora se verifica con facilidad que $e^{0.07862} = 1.0817$. Lo cual significa que, para un incremento unitario en el ingreso ponderado, las posibilidades (ponderadas) en favor de ser propietario de una casa aumentan en 1.0817, o alrededor de 8.17%. *En general, si se toma el antilogaritmo del coeficiente de la j-ésima pendiente (en caso de que haya más de una regresada en el modelo), se resta uno de este valor y se multiplica el resultado por 100, se obtendrá el cambio porcentual en las posibilidades en favor por una unidad de incremento en la j-ésima regresora.*

Por cierto, si desea llevar a cabo el análisis en términos del logit no ponderado, sólo necesita dividir el $L_i^*$ estimado entre $\sqrt{w_i}$. La tabla 15.6 proporciona el logit ponderado estimado y el no ponderado para cada observación, así como algunos otros datos que analizaremos en breve.

### *Cálculo de probabilidades*

Como el lenguaje del logit y de la razón de probabilidades puede ser extraño para algunos, siempre podemos calcular la probabilidad de que una familia con cierto nivel de ingreso posea una casa propia. Suponga que deseamos calcular esta probabilidad para $X = 20$ (20 000). Al insertar este valor en (15.7.1), obtenemos $\hat{L}_i^* = -0.09311$, y al dividir lo anterior entre $\sqrt{w_i} = 4.1816$ (véase la tabla 15.5), obtenemos $\hat{L}_i = -0.02226$. En consecuencia, con un nivel de ingreso de \$20 000, tenemos

$$-0.02199 = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$$

Por tanto,

$$\frac{\hat{P}}{1-\hat{P}_i} = e^{-0.02199} = 0.97825$$

Al resolver para

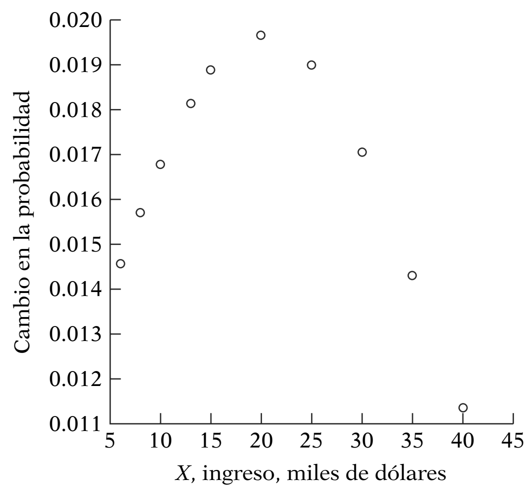
$$\hat{P}_i = \frac{e^{-0.02199}}{1+e^{-0.02199}}$$

**TABLA 15.6**
**Lstar, Xstar, Lstar estimada, probabilidad y cambio en la probabilidad***

| Lstar | Xstar | ELstar | Logit | Probabilidad, $\hat{P}$ | Cambio en la probabilidad† |
|---|---|---|---|---|---|
| −3.50710 | 15.1788 | −2.84096 | −1.12299 | 0.24545 | 0.01456 |
| −3.48070 | 24.15920 | −2.91648 | −0.96575 | 0.27572 | 0.01570 |
| −3.48070 | 35.49600 | −2.86988 | −0.80850 | 0.30821 | 0.01676 |
| −2.64070 | 55.45930 | −2.44293 | −0.57263 | 0.36063 | 0.01813 |
| −0.99850 | 74.62350 | −2.06652 | −0.41538 | 0.39762 | 0.01883 |
| 0.16730 | 83.65060 | −0.09311 | −0.02226 | 0.49443 | 0.01965 |
| 1.60120 | 98.74250 | 1.46472 | 0.37984 | 0.59166 | 0.01899 |
| 2.22118 | 100.48800 | 2.55896 | 0.76396 | 0.68221 | 0.01704 |
| 3.00860 | 95.84050 | 3.16794 | 1.15677 | 0.76074 | 0.01431 |
| 2.77260 | 80.00000 | 3.10038 | 1.55019 | 0.82494 | 0.01135 |

* Lstar y Xstar se tomaron de la tabla 15.5. ELstar es Lstar estimado. Logit es el logit no ponderado. Probabilidad es la probabilidad estimada de ser propietario de una casa. Cambio en la probabilidad es el cambio por unidad de cambio en el ingreso.
† Calculado con base en $\hat{\beta}_2\hat{P}(1-\hat{P}) = 0.07862\,\hat{P}(1-\hat{P})$.

**FIGURA 15.3**
Cambio en la probabilidad en relación con el ingreso.




el lector puede percatarse de que la probabilidad estimada es 0.4945. Es decir, con el ingreso de \$20 000, la probabilidad de que una familia sea dueña de una casa es de casi 49%. La tabla 15.6 muestra las probabilidades así calculadas para diversos niveles de ingreso. Como se ve en dicha tabla, la probabilidad de que una familia tenga casa propia se incrementa, pero no de manera lineal como en el modelo MLP.

### *Cálculo de la tasa de cambio de la probabilidad*

Como se corrobora en la tabla 15.6, la probabilidad de tener una casa propia depende del nivel de ingreso. ¿Cómo podemos calcular la tasa de cambio de las probabilidades conforme varía el ingreso? Como vimos en la nota 19, eso depende no sólo de la pendiente estimada del coeficiente $\beta_2$, sino también del nivel de la probabilidad con que se mida el cambio; desde luego, esto último depende del ingreso con que se calcula la probabilidad.

Para ilustrar lo anterior, suponga que deseamos medir el cambio en la probabilidad de ser propietario de una casa con un nivel de ingreso de \$20 000. Así, de acuerdo con la nota 19, el cambio en la probabilidad correspondiente al incremento en una unidad del ingreso para el nivel 20 (miles) es: $\hat{\beta}(1-\hat{P})\hat{P} = 0.07862(0.5056)(0.4944) = 0.01965$.

Como ejercicio, el lector demostrará que, con un nivel de ingreso de \$40 000, el cambio en la probabilidad es de 0.01135. La tabla 15.6 suministra el cambio en la probabilidad de ser propietario de una casa con diversos niveles de ingreso; dichas probabilidades también se ilustran en la figura 15.3.

Para concluir el análisis de los modelos logit, presentamos a continuación los resultados de la regresión basados en MCO, o regresión no ponderada, para el ejemplo de propiedad de vivienda:

$$\begin{aligned} \hat{L}_i &= \quad -1.6587 \quad + \quad 0.0792X_i \\ \text{ee} &= \quad (0.0958) \qquad (0.0041) \\ t &= (-17.32) \qquad (19.11) \qquad r^2 = 0.9786 \end{aligned} \tag{15.7.3}$$

Se deja al lector la tarea de comparar esta regresión con la de mínimos cuadrados ponderados presentada en la ecuación (15.7.1).

## 15.8 El modelo logit para datos no agrupados o individuales

A fin de establecer el contexto, considere los datos de la tabla 15.7. Sea $Y = 1$, si la calificación final de un estudiante en un curso intermedio de microeconomía fue A, y $Y = 0$ si esa calificación final fue B o C. Spector y Mazzeo utilizaron el GPA (promedio de puntos de calificación),

**TABLA 15.7 Datos sobre el efecto del Sistema de Enseñanza Personalizada (PSI, por sus sigas en inglés) sobre las calificaciones**

| Observación | Calificación GPA | Calificación TUCE | PSI | Calificación | Calificación con letra | Observación | Calificación GPA | Calificación TUCE | PSI | Calificación | Calificación con letra |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.66 | 20 | 0 | 0 | C | 17 | 2.75 | 25 | 0 | 0 | C |
| 2 | 2.89 | 22 | 0 | 0 | B | 18 | 2.83 | 19 | 0 | 0 | C |
| 3 | 3.28 | 24 | 0 | 0 | B | 19 | 3.12 | 23 | 1 | 0 | B |
| 4 | 2.92 | 12 | 0 | 0 | B | 20 | 3.16 | 25 | 1 | 1 | A |
| 5 | 4.00 | 21 | 0 | 1 | A | 21 | 2.06 | 22 | 1 | 0 | C |
| 6 | 2.86 | 17 | 0 | 0 | B | 22 | 3.62 | 28 | 1 | 1 | A |
| 7 | 2.76 | 17 | 0 | 0 | B | 23 | 2.89 | 14 | 1 | 0 | C |
| 8 | 2.87 | 21 | 0 | 0 | B | 24 | 3.51 | 26 | 1 | 0 | B |
| 9 | 3.03 | 25 | 0 | 0 | C | 25 | 3.54 | 24 | 1 | 1 | A |
| 10 | 3.92 | 29 | 0 | 1 | A | 26 | 2.83 | 27 | 1 | 1 | A |
| 11 | 2.63 | 20 | 0 | 0 | C | 27 | 3.39 | 17 | 1 | 1 | A |
| 12 | 3.32 | 23 | 0 | 0 | B | 28 | 2.67 | 24 | 1 | 0 | B |
| 13 | 3.57 | 23 | 0 | 0 | B | 29 | 3.65 | 21 | 1 | 1 | A |
| 14 | 3.26 | 25 | 0 | 1 | A | 30 | 4.00 | 23 | 1 | 1 | A |
| 15 | 3.53 | 26 | 0 | 0 | B | 31 | 3.10 | 21 | 1 | 0 | C |
| 16 | 2.74 | 19 | 0 | 0 | B | 32 | 2.39 | 19 | 1 | 1 | A |

*Notas:* Calificación $Y = 1$ si la calificación final fue A
$Y = 0$ si la calificación final fue B o C
TUCE = calificación en un examen presentado al comienzo del curso para evaluar los conocimientos previos de macroeconomía
PSI = 1 con el nuevo método de enseñanza
= 0 en otro caso
GPA = promedio de puntos de calificación inicial
Fuente: L. Spector y M. Mazzero, "Probit Analysis and Economic Education", *Journal of Economic Education*, vol. 11, 1980, pp. 37-44.

TUCE y PSI (Sistema de Enseñanza Personalizada) de Estados Unidos como predictores de la calificación. El modelo logit en este caso se expresa como:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 \text{GPA}_i + \beta_3 \text{TUCE}_i + \beta_4 \text{PSI}_i + u_i \qquad \textbf{(15.8.1)}$$

Como vimos en la sección 15.6, no podemos poner simplemente $P_i = 1$ si una familia es propietaria de una casa y cero si no lo es. En esta situación, no resultan útiles los MCO ni los MCP. Tenemos que recurrir a procedimientos de cálculo no lineales con el método de máxima verosimilitud. Los detalles de tal método se proporcionan en el apéndice 15A, sección 15A.1. En vista de que el software estadístico más moderno cuenta con rutinas para estimar los modelos logit con base en datos no agrupados, presentaremos los resultados del modelo (15.8.1), que utiliza los datos de la tabla 15.7, y mostraremos la manera de interpretar dichos resultados, mismos que se presentan en la tabla 15.8 en forma tabular. Obtuvimos estos resultados con *EViews 6*. Antes de interpretarlos debemos tener en cuenta las siguientes observaciones generales:

1. Como empleamos el método de máxima verosimilitud, que en general es para muestras grandes, los errores estándar estimados son *asintóticos*.

2. Como resultado, en vez del estadístico $t$ para evaluar la importancia estadística de un coeficiente, empleamos el estadístico (normal estandarizado) $Z$, por lo que las inferencias se basan en la tabla normal. Recuerde que si el tamaño de la muestra es razonablemente grande, la distribución $t$ converge a la distribución normal.

3. Como ya mencionamos, la medida convencional de la bondad de ajuste, $R^2$, no es particularmente significativa para los modelos con regresada binaria. Existen diversas medidas similares

**TABLA 15.8**
**Resultados de la regresión de la ecuación (15.8.1)**

Variable dependiente: Calificación
Método: Logit binario MV
Convergencia lograda después de 5 iteraciones

| Variable | Coeficiente | Error estándar | Estadístico $Z$ | Probabilidad |
|---|---|---|---|---|
| $C$ | -13.0213 | 4.931 | -2.6405 | 0.0082 |
| GPA | 2.8261 | 1.2629 | 2.2377 | 0.0252 |
| TUCE | 0.0951 | 0.1415 | 0.67223 | 0.5014 |
| PSI | 2.3786 | 1.0645 | 2.2345 | 0.0255 |

$R^2$ McFadden = 0.3740    Estadístico RV (3 gl) = 15.40419

a $R^2$, llamadas **pseudo $R^2$**.[26] *EViews* presenta una de esas medidas, la $R^2$ McFadden, denotada como $R^2_{\text{McF}}$, cuyo valor para el ejemplo presente es de 0.3740.[27] Al igual que $R^2$, $R^2_{\text{McF}}$ también varía entre 0 y 1. Otra medida de ajuste comparativamente simple es la **cuenta $R^2$**, que se define como:

$$\textbf{Cuenta } R^2 = \frac{\text{número de predicciones correctas}}{\text{número total de observaciones}} \tag{15.8.2}$$

Como la regresada en el modelo logit toma el valor de 1 o de 0, si la probabilidad pronosticada es mayor que 0.5, se clasifica como si fuese 1, pero si es menor que dicho valor, se considera 0. Así, se cuenta el número de predicciones correctas y se calcula $R^2$ como la dada en (15.8.2). En breve ilustraremos esto.

Debe notarse que, sin embargo, en los modelos con regresada binaria, la bondad del ajuste tiene una importancia secundaria. Lo que interesa son los signos esperados de los coeficientes de la regresión y su importancia práctica y/o estadística.

4. A fin de probar la hipótesis nula respecto de que todos los coeficientes de pendiente son simultáneamente iguales a cero, el equivalente de la prueba $F$ en el modelo de regresión lineal es el **estadístico de la razón de verosimilitud (RV)**. Con la hipótesis nula, el estadístico RV sigue la distribución $\chi^2$ con gl igual al número de variables explicativas, tres para el presente ejemplo. (*Nota:* Excluya el término del intercepto para el cálculo de los gl.)

Ahora interpretemos los resultados de la regresión dados en (15.8.1). En esta ecuación, cada coeficiente de pendiente es un coeficiente de *pendiente parcial* y mide el cambio en el logit estimado correspondiente a una unidad de cambio del valor de la regresada dada (con las demás regresoras constantes). Por tanto, el coeficiente del GPA igual a 2.8261 significa que, mientras las demás variables se mantengan constantes, si el GPA se incrementa en una unidad, en promedio el logit estimado aumenta casi 2.83 unidades, lo cual indica una relación positiva entre ambos. Como se aprecia, todas las demás regresoras tienen un efecto positivo en el logit, a pesar de que en términos estadísticos el efecto de TUCE no es importante. No obstante, todas las regresoras en conjunto tienen un impacto importante en la calificación final, pues el estadístico RV es igual a 15.40, cuyo valor $p$ es de casi 0.0015, el cual resulta muy pequeño.

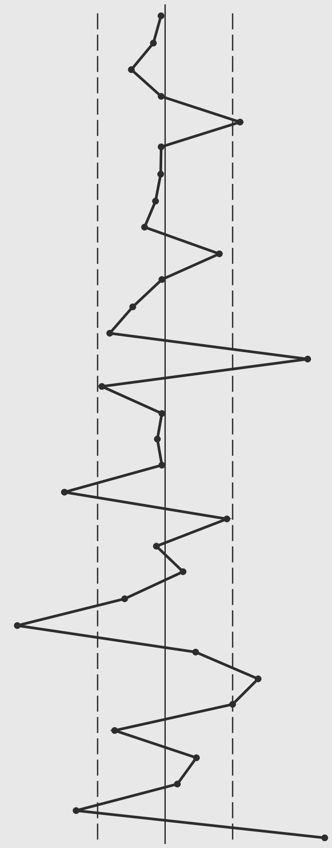
Como dijimos antes, una interpretación más significativa se da en términos de las posibilidades en favor, las cuales se obtienen al tomar el antilogaritmo de los diversos coeficientes de pendiente. En consecuencia, si tomamos el antilogaritmo del coeficiente de PSI, igual a 2.3786,

[26] Para un análisis comprensible, véase J. Scout Long, *Regression Models for Categorical and Limited Dependent Variables*, Sage, Newbury Park, California, 1997, pp. 102-113.

[27] Técnicamente, esto se define como: $1 - (\text{FLV}_{\text{ir}}/\text{FLV}_r)$, donde $\text{FLV}_{\text{ir}}$ es la función logaritmo de verosimilitud irrestricta, para la cual se incluyen en el modelo todas las regresoras, y $\text{FLV}_r$ es la función logaritmo de verosimilitud restringida para la cual sólo se incluye el intercepto en el modelo. Conceptualmente, $\text{FLV}_{\text{ir}}$ es el equivalente de la SCR y $\text{FLV}_r$ es el equivalente de la SCT del modelo de regresión lineal.

**TABLA 15.9**
**Valores reales y ajustados con base en la regresión de la tabla 15.8**

| Observación | Real | Ajustado | Residuo | Gráfico de residuos |
|---|---|---|---|---|
| 1 | 0 | 0.02658 | −0.02658 | |
| 2 | 0 | 0.05950 | −0.05950 | |
| 3 | 0 | 0.18726 | −0.18726 | |
| 4 | 0 | 0.02590 | −0.02590 | |
| 5 | 1 | 0.56989 | 0.43011 | |
| 6 | 0 | 0.03486 | −0.03486 | |
| 7 | 0 | 0.02650 | −0.02650 | |
| 8 | 0 | 0.05156 | −0.05156 | |
| 9 | 0 | 0.11113 | −0.11113 | |
| 10 | 1 | 0.69351 | 0.30649 | |
| 11 | 0 | 0.02447 | −0.02447 | |
| 12 | 0 | 0.19000 | −0.19000 | |
| 13 | 0 | 0.32224 | −0.32224 | |
| *14 | 1 | 0.19321 | 0.80679 | |
| 15 | 0 | 0.36099 | −0.36099 | |
| 16 | 0 | 0.03018 | −0.03018 | |
| 17 | 0 | 0.05363 | −0.05363 | |
| 18 | 0 | 0.03859 | −0.03859 | |
| *19 | 0 | 0.58987 | −0.58987 | |
| 20 | 1 | 0.66079 | 0.33921 | |
| 21 | 0 | 0.06138 | −0.06138 | |
| 22 | 1 | 0.90485 | 0.09515 | |
| 23 | 0 | 0.24177 | −0.24177 | |
| *24 | 0 | 0.85209 | −0.85209 | |
| 25 | 1 | 0.83829 | 0.16171 | |
| *26 | 1 | 0.48113 | 0.51887 | |
| 27 | 1 | 0.63542 | 0.36458 | |
| 28 | 0 | 0.30722 | −0.30722 | |
| 29 | 1 | 0.84170 | 0.15830 | |
| 30 | 1 | 0.94534 | 0.05466 | |
| *31 | 0 | 0.52912 | −0.52912 | |
| *32 | 1 | 0.11103 | 0.88897 | |



* Predicciones incorrectas.

obtendremos 10.7897 ($\approx e^{2.3786}$). Esto indica que los estudiantes expuestos al nuevo método de enseñanza son por encima de 10 veces más propensos a obtener una A que quienes no están expuestos al nuevo método, en tanto no cambien los demás factores.

Suponga que deseamos calcular la probabilidad real de que un estudiante obtenga una calificación de A. Considere al estudiante 10 de la tabla 15.7. Al introducir los datos reales de este estudiante en el modelo logit estimado, dado en la tabla 15.8, puede verificar que el valor logit estimado para dicho estudiante es igual a 0.8178. Con la ecuación (15.5.2) puede verificar con facilidad que la probabilidad estimada es 0.69351. Como la calificación final real del estudiante fue A y el modelo logit asigna una probabilidad de 1 a un estudiante que obtiene una A, la probabilidad estimada de 0.69351 no es exactamente 1, pero se le aproxima.

Recuerde la cuenta $R^2$ ya definida. La tabla 15.9 proporciona los valores reales y pronosticados de la regresada para el ejemplo ilustrativo. En esta tabla se observa que, de las 32 observaciones, hubo 6 pronósticos incorrectos (estudiantes 14, 19, 24, 26, 31 y 32). Por tanto, el valor de la cuenta $R^2$ es $26/32 = 0.8125$, en tanto que el valor $R^2$ McFadden es 0.3740. Aunque estos dos valores no se pueden comparar de manera directa, dan una idea de los órdenes de magnitud. Además, no se debe sobrevalorar la importancia de la bondad de ajuste en modelos para los que la regresada es dicótoma.

**EJEMPLO 15.5**
*¿Quién tiene tarjeta de débito? Análisis logit*

Ya estudiamos los resultados del modelo lineal de probabilidad (MLP) aplicado a los datos de tarjetas de débito bancarias, por lo que ahora aplicaremos el modelo logit. Los resultados son como sigue:

```
Variable dependiente: DÉBITO
Método: Logit binario MV (ascenso de colina cuadrático)
Muestra: 1-60
Observaciones incluidas: 60
Convergencia lograda después de 4 iteraciones
Matriz de covarianza calculada con segundas derivadas
```

| Variable | Coeficiente | Error estándar | Estadístico *z* | Probabilidad |
|---|---|---|---|---|
| C | -0.574900 | 0.785787 | -0.731624 | 0.4644 |
| Saldo | 0.001248 | 0.000697 | 1.789897 | 0.0735 |
| Cajero automático | -0.120225 | 0.093984 | -1.279205 | 0.2008 |
| Interés | -1.352086 | 0.680988 | -1.985478 | 0.0471 |

| | | | |
|---|---|---|---|
| *R* cuadrada McFadden | 0.080471 | Media de la variable dependiente | 0.433333 |
| Desviación estándar de la variable dependiente | 0.499717 | Error estándar de la regresión | 0.486274 |
| | | Suma de cuadrados residual | 13.24192 |
| Criterio de información de Akaike | 1.391675 | Log verosimilitud | -37.75024 |
| Criterio de Schwarz | 1.531298 | Log verosimilitud restr. | -41.05391 |
| Criterio de Hannan-Quinn | 1.446289 | Promedio log verosimilitud | -0.629171 |
| Estadístico RV | 6.607325 | | |
| Probabilidad (estadístico RV) | 0.085525 | | |
| Obs. con Dep = 0 | 34 | Total de obs. | 60 |
| Obs. con Dep = 1 | 26 | | |

El signo positivo de Saldo y los signos negativos de Cajero automático e Interés se asemejan a los del MLP, aunque no podemos establecer una comparación directa entre los dos. La interpretación de los coeficientes en el modelo logit difiere de la del MLP. En este caso, por ejemplo, si la tasa de interés aumenta un punto porcentual, el logit disminuye alrededor de 1.35, si las demás variables permanecen constantes. Si tomamos el antilogaritmo de −1.352086, obtenemos casi 0.2587. Esto significa que si se paga una tasa de interés sobre los saldos de las cuentas, es probable que, en promedio, sólo alrededor de uno de cada cuatro clientes tenga una tarjeta de débito.

Con base en el estadístico RV estimado, observamos que, en conjunto, las tres variables son estadísticamente significativas en el nivel aproximado de 8.5%. Si usamos el nivel convencional de significancia de 5%, estas variables son sólo marginalmente significativas.

El valor $R^2$ McFadden es muy bajo. Con los datos, el lector podrá averiguar el valor de la cuenta $R^2$.

Como ya señalamos, a diferencia del MLP, los coeficientes de pendiente no dan la tasa de cambio de la probabilidad por cada unidad de cambio en la regresora. Es necesario calcularlos como se muestra en la tabla 15.6. Por fortuna, esta tarea manual no es necesaria, pues hay programas estadísticos, como el STATA, que lo hacen de forma rutinaria. En este ejemplo los resultados son los siguientes:

Efectos marginales después de logit:

$$Y = \text{Pr(débito)(pronóstico)}$$
$$= .42512423$$

| Variable | dy/dx | Error estándar | *z* | *p* > \|*z*\| | [ 95% C. I. | ] | *x* |
|---|---|---|---|---|---|---|---|
| Saldo | .000305 | .00017 | 1.79 | 0.073 | -.000029 | .000639 | 1 499.87 |
| Interés* | -.2993972 | .12919 | -2.32 | 0.020 | -.552595 | -.046199 | .266667 |
| Cajero automático | -.0293822 | .02297 | -1.28 | 0.201 | -.074396 | .015631 | 10.3 |

* dy/dx representa el cambio discreto de la variable dicótoma de 0 a 1.

(*continúa*)

**EJEMPLO 15.5** (*continuación*)

El coeficiente 0.000305 indica que los clientes que mantienen saldos altos tienen 0.03% más probabilidades de tener una tarjeta de débito, pero si la tasa de interés aumenta un punto porcentual, la probabilidad de tener una tarjeta de débito disminuye casi 30%. El coeficiente de Cajero automático, aunque estadísticamente insignificante, indica que si las transacciones en cajero automático aumentan una unidad, la probabilidad de tener una tarjeta de débito disminuye casi 2.9 por ciento.

## 15.9 Modelo probit

Como mencionamos, para explicar el comportamiento de una variable dependiente dicótoma es preciso utilizar una función de distribución acumulativa (FDA) seleccionada apropiadamente. El modelo logit utiliza la función logística acumulativa, como se indica en (15.5.2), pero no es la única FDA posible. En algunas aplicaciones, la FDA normal se ha encontrado útil. El modelo de estimación que surge de una FDA[28] normal se conoce comúnmente como **modelo probit**, aunque algunas veces también como **modelo normit**. En principio, se puede sustituir la FDA normal por la FDA logística en (15.5.2) y proceder como en la sección 16.5. Pero en lugar de seguir este camino, presentaremos el modelo probit basado en la teoría de la utilidad, o de la perspectiva de selección racional con base en el comportamiento, según el modelo desarrollado por McFadden.[29]

Para motivar el modelo probit, suponga que en el ejemplo de propiedad de vivienda, la decisión de la $i$-ésima familia de tener casa propia o de no tenerla depende de un **índice de conveniencia** *no observable* $I_i$ (también conocido como **variable latente**), determinado por una o diversas variables explicativas, digamos, el ingreso $X_i$, de manera que entre mayor sea el valor del índice $I_i$, mayor será la probabilidad de que la familia posea vivienda. Expresamos el índice $I_i$ como

$$I_i = \beta_1 + \beta_2 X_i \tag{15.9.1}$$

donde $X_i$ es el ingreso de la $i$-ésima familia.

¿Cómo se relaciona el índice (no observable) con la decisión real de ser propietario de una casa? Igual que antes, sea $Y = 1$ si una familia es propietaria de una casa y $Y = 0$ si no lo es. Ahora bien, es razonable suponer que para cada familia hay un **nivel crítico** o **umbral del índice**, que podemos denominar $I_i^*$, tal que si $I_i$ excede a $I_i^*$, la familia tendrá una casa propia, de lo contrario no lo hará. El umbral $I_i^*$, al igual que $I_i$, no es observable, pero si suponemos que está distribuido normalmente con las mismas media y varianza, es posible no sólo estimar los parámetros del índice dado en (15.9.1), sino también obtener alguna información sobre el propio índice no observable. Este cálculo se muestra de la siguiente manera.

Con el supuesto de normalidad, la probabilidad de que $I_i^*$ sea menor o igual que $I_i$ se calcula a partir de la FDA normal estándar como:[30]

$$P_i = P(Y = 1 \mid X) = P(I_i^* \le I_i) = P(Z_i \le \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i) \tag{15.9.2}$$

[28] Véase el **apéndice A** para un análisis de la FDA normal. Brevemente, si una variable $X$ sigue la distribución normal con media $\mu$ y varianza $\sigma^2$, su FDP es

$$f(X) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

y su FDA es

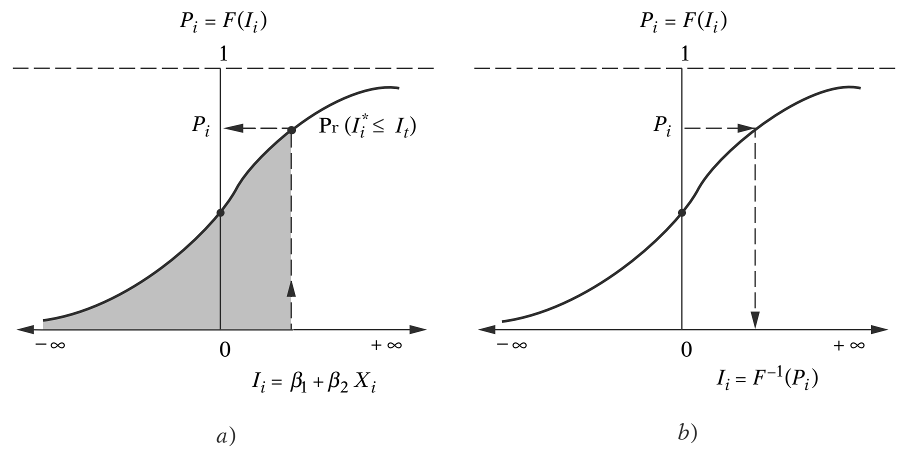
$$F(X) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

donde $X_0$ es algún valor especificado de $X$.

[29] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behaviour", en P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, Nueva York, 1973.

[30] Una distribución normal con media cero y varianza unitaria (= 1) se conoce como variable normal estándar o estandarizada. (Véase el **apéndice A**.)

**FIGURA 15.4**
Modelo probit: *a*) dado $I_i$, lea $P_i$ en la ordenada; *b*) dado $P_i$, lea $I_i$ en la abscisa.




donde $P(Y = 1 \mid X)$ significa la probabilidad de que un suceso ocurra dado(s) el (los) valor(es) de $X$ o la(s) variable(s) explicativa(s), y donde $Z_i$ es la variable normal estandarizada; es decir, $Z \sim N(0, \sigma^2)$. $F$ es la FDA normal estándar, que, escrita de manera explícita en el presente contexto, es:

$$
\begin{aligned}
F(I_i) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2}\, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-z^2/2}\, dz
\end{aligned}
\tag{15.9.3}
$$

Como $P$ representa la probabilidad de que ocurra un suceso, en este caso la probabilidad de tener casa propia, se mide por el área de la curva normal estándar de $-\infty$ a $I_i$, como se muestra en la figura 15.4*a*.

Ahora, para obtener información sobre $I_i$, el índice de utilidad, lo mismo que para $\beta_1$ y $\beta_2$, se toma la inversa de (15.9.2) para obtener

$$
\begin{aligned}
I_i &= F^{-1}(I_i) = F^{-1}(P_i) \\
&= \beta_1 + \beta_2 X_i
\end{aligned}
\tag{15.9.4}
$$

donde $F^{-1}$ es la inversa de la FDA normal. El significado de todo esto se aclara con la figura 15.4. En el panel *a*) de esta figura se obtiene (de la ordenada) la probabilidad (acumulada) de tener casa propia dado $I_i^* \leq I_i$, mientras que en el panel *b*) se obtiene (de la abscisa) el valor de $I_i$ dado el valor de $P_i$, que es simplemente el inverso del primero.

Pero, específicamente, ¿cómo obtenemos el índice $I_i$ al igual que las estimaciones de $\beta_1$ y $\beta_2$? Como en el caso del modelo logit, la respuesta depende que se cuente con datos agrupados o desagrupados. Consideramos los dos casos de manera individual.

## Estimación de probit con datos agrupados: gprobit

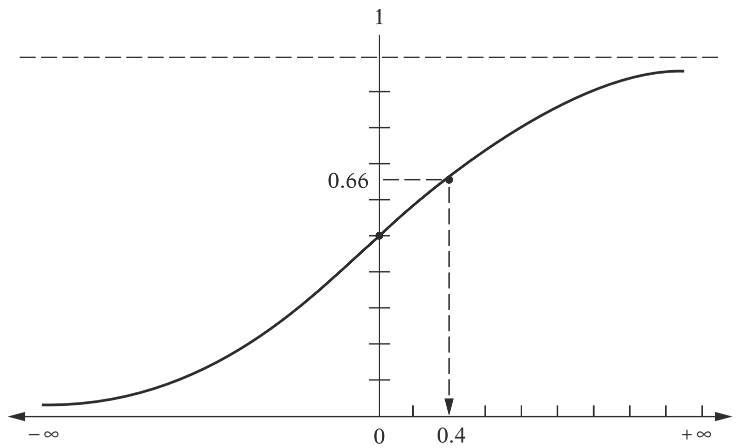
Utilizaremos los mismos datos que para glogit, los cuales se muestran en la tabla 15.4. Como ya tenemos $\hat{P}_i$, la frecuencia relativa (la medida empírica de probabilidad) de poseer una casa propia para diversos niveles de ingresos, como se muestra en la tabla 15.5, con ella obtenemos $I_i$ de la FDA normal, como se muestra en la tabla 15.10, o en la figura 15.5.

**TABLA 15.10**
**Estimación del índice $I_i$ de la FDA normal estándar**

| $\hat{P}_i$ | $I_i = F^{-1}(\hat{P}_i)$ |
|---|---|
| 0.20 | −0.8416 |
| 0.24 | −0.7063 |
| 0.30 | −0.5244 |
| 0.35 | −0.3853 |
| 0.45 | −0.1257 |
| 0.51 | 0.0251 |
| 0.60 | 0.2533 |
| 0.66 | 0.4125 |
| 0.75 | 0.6745 |
| 0.80 | 0.8416 |

*Notas:* 1) Los $\hat{P}_i$ provienen de la tabla 15.5; 2) los $I_i$ se calculan de la FDA normal estándar.

**FIGURA 15.5**
FDA normal.




Una vez estimado el $I_i$, estimar $\beta_1$ y $\beta_2$ es un asunto relativamente sencillo, como mostraremos en breve. A propósito, observe que en el lenguaje del análisis probit, el índice de utilidad no observable $I_i$ se conoce simplemente como **desviación equivalente normal** (d.e.n.) o sólo **normit**. Como la d.e.n. o $I_i$ será negativa siempre que $P_i < 0.5$, en la práctica se agrega el número 5 a la d.e.n. y el resultado se denomina un probit.

**EJEMPLO 15.6**
*Ilustración de gprobit mediante el ejemplo de propiedad de una casa*

Continuemos con el ejemplo de propiedad de una casa. Ya presentamos los resultados de los modelos glogit para este ejemplo. Los resultados del probit agrupado (gprobit) de los mismos datos son los siguientes:

Con la d.e.n. (= $I$) dada en la tabla 15.10, los resultados de la regresión son los de la tabla 15.11.[31] Los resultados de la regresión basados en los probits (= d.e.n. + 5) se dan en la tabla 15.12.

A excepción del término del intercepto, estos resultados son idénticos a los de la tabla anterior. Pero esto no debe sorprender. (¿Por qué?)

[31] Los siguientes resultados no están corregidos por heteroscedasticidad. En el ejercicio 15.12 se propone un procedimiento apropiado con el objeto de corregir la heteroscedasticidad.

**EJEMPLO 15.6** (*continuación*)

**TABLA 15.11**

Variable dependiente: *I*

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C* | -1.0166 | 0.0572 | -17.7473 | 1.0397E-07 |
| Ingreso | 0.04846 | 0.00247 | 19.5585 | 4.8547E-08 |

$R^2$ = 0.97951 Estadístico de Durbin-Watson = 0.91384

**TABLA 15.12**

Variable dependiente: Probit

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C* | 3.9833 | 0.05728 | 69.5336 | 2.03737E-12 |
| Ingreso | 0.04846 | 0.00247 | 19.5585 | 4.8547E-08 |

$R^2$ = 0.9795 Estadístico de Durbin-Watson = 0.9138

*Nota:* Estos resultados no están corregidos por heteroscedasticidad (véase el ejercicio 15.12).

### *Interpretación de las estimaciones de probit en la tabla 15.11*

¿Cómo interpretamos los resultados anteriores? Suponga que deseamos averiguar el efecto de una unidad de cambio en $X$ (ingreso medido en miles de dólares) sobre la probabilidad de que $Y = 1$; es decir, que una familia adquiera una casa. Para hacerlo, examine la ecuación (15.9.2). Deseamos tomar la derivada de esta función respecto de $X$ (es decir, la tasa de cambio de la probabilidad respecto del ingreso). Resulta que su derivada es:

$$\frac{dP_i}{dX_i} = f(\beta_1 + \beta_2 X_i)\beta_2 \tag{15.9.5}$$ [32]

donde $f(\beta_1 + \beta_2 X_i)$ es la función de densidad de probabilidad de la normal estandarizada evaluada en $\beta_1 + \beta_2 X_i$. Resulta evidente que esta evaluación dependerá del valor particular de las variables $X$. Tome el valor de $X$ de la tabla 15.5, por ejemplo, $X = 6$ (miles de dólares). Con los valores estimados de los parámetros proporcionados en la tabla 15.11, deseamos encontrar la función de densidad normal en $f[-1.0166 + 0.04846(6)] = f(-0.72548)$. Si consulta las tablas de distribución normal, encontrará que para $Z = -0.72548$, la densidad normal es casi de 0.3066.[33] Ahora, si multiplica ese valor por el coeficiente de pendiente estimado de 0.04846, obtiene 0.01485. Esto significa que si empieza con un nivel de ingreso de $6 000 y el ingreso aumenta $1 000, la probabilidad de que una familia compre una casa se incrementa alrededor de 1.4%. (Compare este resultado con el de la tabla 15.6.)

Como se ve en el análisis anterior, el cálculo de cambios en la probabilidad mediante el modelo probit resulta un poco tedioso, en comparación con los modelos MLP y logit.

En vez de calcular los cambios en la probabilidad, suponga que deseamos encontrar las probabilidades estimadas a partir del modelo gprobit ajustado. Esto se realiza con facilidad. Con los

[32] Utilizamos la regla de la cadena de las derivadas:

$$\frac{dP_i}{dX_i} = \frac{dF(t)}{dt} \cdot \frac{dt}{dX}$$

donde $t = \beta_1 + \beta_2 X_i$.

[33] Observe que la normal estandarizada $Z$ puede variar de $-\infty$ hasta $+\infty$, pero la función de densidad $f(Z)$ siempre es positiva.

datos de la tabla 15.11 y al intercalar los valores de *X* de la tabla 15.5, el lector puede verificar que los valores estimados n.i.d. (hasta dos dígitos) son los siguientes:

| *X* | 6 | 8 | 10 | 13 | 15 | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| n.i.d. estimados | −0.72 | −0.63 | −0.53 | −0.39 | −0.29 | −0.05 | 0.19 | 0.43 | 0.68 | 0.92 |

En la actualidad, los programas estadísticos como MINITAB calculan fácilmente las probabilidades (acumulativas) asociadas a los diversos n.i.d. Por ejemplo, la probabilidad estimada correspondiente a un valor n.i.d. de −0.63 es igual a 0.2647, y la correspondiente a un valor n.i.d. de 0.43 es de 0.6691. Si comparamos estas estimaciones con los valores reales dados en la tabla 15.5, encontrará que ambos están muy cercanos entre sí, lo cual indica que el modelo ajustado es bastante bueno. Gráficamente, lo que acabamos de realizar ya se presentó en la figura 15.4.

## El modelo probit para datos no agrupados o individuales

Vuelva a examinar la tabla 15.7, que proporciona los datos de 32 individuos respecto de sus calificaciones finales en los exámenes de microeconomía intermedia, en relación con las variables GPA, TUCE y PSI. Los resultados de la regresión logit se presentan en la tabla 15.8. Vea cómo lucen los resultados probit. Observe que, como en el caso del modelo logit para datos individuales, tendremos que utilizar un procedimiento de cálculo no lineal basado en el método de máxima verosimilitud. Los resultados de la regresión calculados mediante *EViews 6* se dan en la tabla 15.13.

"Cualitativamente", los resultados del modelo probit se pueden comparar con los obtenidos mediante el modelo logit, pues tanto GPA como PSI son estadísticamente significativas en lo individual. En conjunto, todos los coeficientes son estadísticamente significativos, pues el valor del estadístico RV es 15.5458, con un valor *p* de 0.0014. Por razones dadas en secciones posteriores, no podemos comparar de manera directa los coeficientes de regresión logit y probit.

Para propósitos comparativos, presentamos los resultados basados en el modelo lineal de probabilidad (MLP) para los datos de calificaciones de la tabla 15.14. De nuevo, cualitativa-

**TABLA 15.13**

Variable dependiente: Calificación
Método: Probit binario MV
Convergencia lograda después de 5 iteraciones

| Variable | Coeficiente | Error estándar | Estadístico *Z* | Probabilidad |
|---|---|---|---|---|
| *C* | -7.4523 | 2.5424 | -2.9311 | 0.0033 |
| GPA | 1.6258 | 0.6938 | 2.3430 | 0.0191 |
| TUCE | 0.0517 | 0.0838 | 0.6166 | 0.5374 |
| PSI | 1.4263 | 5950 | 2.3970 | 0.0165 |

Estadístico RV (3 gl) = 15.5458 $R^2$ McFadden = 0.3774
Probabilidad (estadístico RV) = 0.0014

**TABLA 15.14**

Variable dependiente: Calificación

| Variable | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C* | -1.4980 | 0.5238 | -2.8594 | 0.0079 |
| GPA | 0.4638 | 0.1619 | 2.8640 | 0.0078 |
| TUCE | 0.0104 | 0.0194 | 0.5386 | 0.5943 |
| PSI | 0.3785 | 0.1391 | 2.7200 | 0.0110 |

$R^2$ = 0.4159 Estadístico *d* de Durbin-Watson = 2.3464
Estadístico *F* = 6.6456

mente, los resultados MLP resultan similares a los de los modelos logit y probit, pues GPA y PSI son estadísticamente significativos en lo individual, pero TUCE no lo es. Asimismo, las variables explicativas en conjunto tienen un gran efecto en la calificación, pues el valor $F$ de 6.6456 es estadísticamente significativo: su valor $p$ es sólo de 0.0015.

### Efecto marginal de un cambio unitario en el valor de una regresora sobre los diversos modelos de regresión

En el *modelo de regresión lineal*, el coeficiente de la pendiente mide el cambio en el valor promedio de la regresada, debido a una unidad de cambio en el valor de la regresora, con las demás variables constantes.

En el *MLP*, el coeficiente de la pendiente mide directamente el cambio en la probabilidad de que ocurra un evento, como resultado de una unidad de cambio en el valor de la regresora, con un efecto constante de todas las demás variables.

En el *modelo logit*, el coeficiente de la pendiente de una variable indica el cambio en el logaritmo de las posibilidades en favor de que ocurra un evento asociadas a una unidad de cambio en esa variable, de nuevo, con todas las demás variables constantes. Pero como ya dijimos, para el modelo logit la tasa de cambio en la probabilidad de que ocurra un suceso está dada por $\beta_j P_i (1 - P_i)$, donde $\beta_j$ es el coeficiente (de regresión parcial) de la $j$-ésima regresora. Pero al evaluar $P_i$, participan todas las variables incluidas en el análisis.

En el *modelo probit*, como ya vimos, la tasa de cambio de la probabilidad es un tanto complicada y está dada por $\beta_j f(Z_i)$, donde $f(Z_i)$ es la función de densidad de la variable normal estandarizada y $Z_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$; es decir, el modelo de regresión utilizado en el análisis.

Por tanto, en los modelos probit y logit todas las regresoras intervienen en el cálculo de los cambios en la probabilidad, en tanto que en el MLP sólo participa la $j$-ésima regresora. Esta diferencia tal vez sea la razón de la anterior popularidad del modelo MLP. Los programas estadísticos como STATA facilitan ahora en gran medida la tarea de calcular la tasa de cambio de la probabilidad en los modelos logit y probit. Por tanto, en la actualidad no hay necesidad de elegir el MLP sólo por su sencillez.

## 15.10 Modelos logit y probit

Aunque para el ejemplo de las calificaciones los modelos logit, probit y MLP dan cualitativamente resultados semejantes, nos centraremos en los modelos logit y probit, en vista de los problemas con el MLP ya mencionados. De los modelos logit y probit, ¿cuál preferiría? Para la mayoría de las aplicaciones, los modelos son muy semejantes; la principal diferencia es que la distribución logística tiene colas un poco más anchas, lo cual se observa en la figura 15.6. Lo anterior significa que la probabilidad condicional $P_i$ se aproxima a cero o a uno con una tasa menor en el modelo logit, en comparación con el probit. Esto se observa con mayor claridad en la tabla 15.15. Por consiguiente, no existe una razón de peso para elegir uno y no otro. En la práctica, muchos investigadores eligen el modelo logit debido a su comparativa simplicidad matemática.

Aunque los modelos son similares, se debe tener cuidado al interpretar los coeficientes estimados por ambos modelos. Por ejemplo, para el caso de las calificaciones, el coeficiente de GPA, igual a 1.6528 en el modelo probit (tabla 15.13) e igual a 2.8261 en el modelo logit (tabla 15.8), no son directamente comparables. La razón es que aunque las distribuciones logística estándar (la base para el logit) y normal estándar (la base para el probit) tienen ambas media cero, sus varianzas son diferentes; 1 para la normal estándar (como ya sabíamos) y $\pi^2/3$ para la distribución logística, donde $\pi \approx 22/7$. En consecuencia, si multiplicamos el coeficiente probit por aproximadamente 1.81 (casi igual a $\pi/\sqrt{3}$), tendremos aproximadamente el coeficiente logit. Para nuestro ejemplo, el coeficiente probit de GPA es 1.6258. Al multiplicarlo por 1.81, obtenemos 2.94, cercano al coeficiente logit. Por otra parte, si multiplicamos un coeficiente logit por 0.55 ($= 1/1.81$),

**FIGURA 15.6**
Distribuciones acumulativas logit y probit.

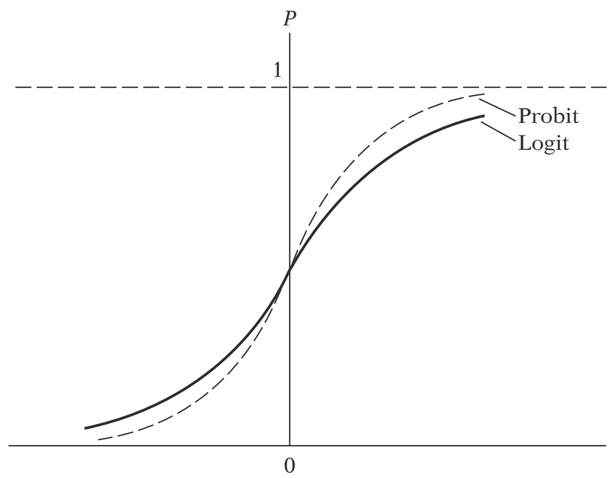



**TABLA 15.15**
**Valores de las funciones de probabilidad acumulativas**

| | Normal acumulativa | Logística acumulativa |
|---|---|---|
| $Z$ | $P_1(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-s^2/2} ds$ | $P_2(Z) = \frac{1}{1 + e^{-z}}$ |
| −3.0 | 0.0013 | 0.0474 |
| −2.0 | 0.0228 | 0.1192 |
| −1.5 | 0.0668 | 0.1824 |
| −1.0 | 0.1587 | 0.2689 |
| −0.5 | 0.3085 | 0.3775 |
| 0 | 0.5000 | 0.5000 |
| 0.5 | 0.6915 | 0.6225 |
| 1.0 | 0.8413 | 0.7311 |
| 1.5 | 0.9332 | 0.8176 |
| 2.0 | 0.9772 | 0.8808 |
| 3.0 | 0.9987 | 0.9526 |

tendremos el coeficiente probit. Sin embargo, Amemiya sugiere multiplicar una estimación logit por 0.625, a fin de obtener una mejor estimación para el correspondiente probit estimado.[34] En sentido inverso, al multiplicar un coeficiente probit por 1.6 (= 1/0.625), obtenemos el coeficiente logit correspondiente.

A propósito, Amemiya también demostró que los coeficientes de los modelos MLP y logit se relacionan de la siguiente manera:

$$\beta_{\text{MLP}} = 0.25\beta_{\text{logit}} \qquad \text{excepto para el intercepto}$$

y

$$\beta_{\text{MLP}} = 0.25\beta_{\text{logit}} + 0.5 \qquad \text{para el intercepto}$$

Se deja al lector la tarea de verificar si las aproximaciones son válidas para el ejemplo de las calificaciones.

Para concluir el análisis de los modelos MLP, logit y probit, consideraremos otro ejemplo.

[34] T. Amemiya, "Qualitative Response Model: A Survey", *Journal of Economic Literature,* vol. 19, 1981, pp. 481-536.

**EJEMPLO 15.7**
*Fumar o no fumar*

Para averiguar qué factores determinan si una persona será fumadora o no, obtuvimos datos sobre 1 196 sujetos de estudio.[35] Para cada uno de ellos se presenta información sobre escolaridad, edad, ingreso y precio de los cigarrillos en 1979. La variable dependiente es fumador, donde 1 corresponde a los fumadores y 0 a los no fumadores. En el ejercicio 15.20 realizaremos otro análisis basado en los datos de la tabla 15.28 del sitio web del libro. Para efectos de comparación, presentamos los resultados basados en los modelos MLP, logit y probit en forma tabular (tabla 15.16). Estos resultados se obtuvieron con la versión 10 de STATA.

**TABLA 15.16**

| Variables | MLP | Logit | Probit |
|---|---|---|---|
| Constante | 1.1230 | 2.7450 | 1.7019 |
| | (5.96) | (3.31) | (3.33) |
| Edad | −0.0047 | −0.0208 | −0.0129 |
| | (−5.70) | (−5.58) | (−5.66) |
| Escolaridad | −0.0206 | −0.0909 | −0.0562 |
| | (−4.47) | (−4.40) | (−4.45) |
| Ingreso | 1.03e−0.6 | 4.72e−06 | 2.72e−06 |
| | (0.63) | (0.66) | (0.62) |
| Pcigs79 | −0.0051 | −0.0223 | −0.0137 |
| | (−1.80) | (−1.79) | (−1.79) |
| $R^2$ | 0.0388 | 0.0297 | 0.0301 |

*Notas:* Las cifras entre paréntesis son razones $t$ en el MLP y $z$ en los modelos logit y probit. En logit y probit, los valores $R^2$ son valores pseudo $R^2$.

Aunque los coeficientes de los tres modelos no se pueden comparar de manera directa, en el sentido cualitativo son similares. Así, edad, escolaridad y precio de los cigarrillos producen efecto negativo en el hábito de fumar, y el ingreso tiene efecto positivo. Estadísticamente, el efecto del ingreso es cero y el efecto del precio es significativo en un nivel aproximado a 8%. En el ejercicio 15.20 se pide al lector que aplique el factor de conversión para producir varios coeficientes comparables.

En la tabla 15.17 se presenta el *efecto marginal* de cada variable sobre la probabilidad de fumar en cada tipo de modelo.

**TABLA 15.17**

| Variables | MLP | Logit | Probit |
|---|---|---|---|
| Edade | −0.0047 | −0.0048 | −0.0049 |
| Escolaridad | −0.0206 | −0.0213 | −0.0213 |
| Ingreso | 1.03e−06 | 1.11e−06 | 1.03e−06 |
| Pcigs79 | −0.0051 | −0.0052 | −0.0052 |

*Nota:* Salvo por el ingreso, los coeficientes estimados son muy significativos estadísticamente para edad y escolaridad, y significativos en un nivel aproximado a 8% para el precio de los cigarrillos.

Como reconocerá, en el MLP el efecto marginal de una variable sobre la probabilidad de fumar se obtiene directamente de los coeficientes de regresión estimados, pero en los modelos logit y probit deben calcularse como explicamos en el capítulo.

Es interesante que los efectos marginales de los tres modelos sean muy parecidos. Por ejemplo, si el nivel de escolaridad aumenta, en promedio, la probabilidad de que alguien se convierta en fumador se reduce en alrededor de 2%.

[35] Estos datos son de Michael P. Murray, *Econometrics: A Modern Introduction,* Pearson/Addison-Wesley, Boston, 2006, y pueden descargarse de www.aw-bc.com/murray.

## 15.11 Modelo tobit

Una extensión del modelo probit es el **modelo tobit**, desarrollado por James Tobin, economista laureado con el Nobel. Para explicar este modelo continuamos con el ejemplo de propiedad de vivienda. En el modelo probit, el objetivo era estimar la probabilidad de tener una casa propia como función de algunas variables socioeconómicas. En el modelo tobit, el interés radica en averiguar la cantidad de dinero que una persona o familia gasta en una casa en relación con las variables socioeconómicas. Ahora tenemos un problema: si un consumidor no compra una casa, obviamente no hay información sobre el gasto en vivienda de tales consumidores; se tiene tal información sólo sobre los consumidores que en efecto compran casa.

Así, los consumidores se dividen en dos grupos, uno que consiste, por ejemplo, en $n_1$ consumidores de quienes se posee información sobre las regresoras (por ejemplo, ingreso, tasa de interés hipotecaria, número de personas que forman la familia, etc.) al igual que sobre la variable regresada (cantidad de gasto en vivienda), y otro que consiste, por ejemplo, en $n_2$ consumidores de quienes sólo se tiene información sobre las regresoras pero no sobre la variable regresada. Cuando en una muestra la información sobre la variable regresada está disponible sólo para algunas observaciones, se conoce como **muestra censurada**.[36] Por consiguiente, el modelo tobit también se conoce como modelo de regresión censurada. Algunos autores los llaman **modelos de regresión con variable dependiente limitada** debido a la restricción impuesta sobre los valores tomados por la variable regresada.

En términos estadísticos, el modelo tobit se expresa como

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i && \text{si LD} > 0 \\ &= 0 && \text{en otro caso} \end{aligned} \tag{15.11.1}$$

donde LD = lado derecho. *Nota:* Se pueden agregar fácilmente otras variables $X$ al modelo.

¿Es posible estimar la regresión (15.11.1) sólo con $n_1$ observaciones y dejar de lado sin más la preocupación por las $n_2$ observaciones restantes? La respuesta es no, pues las estimaciones por MCO de los parámetros obtenidos del subconjunto de $n_1$ observaciones estarán *sesgadas y serán inconsistentes*; es decir, estarán sesgadas pero de manera asintótica.[37]

Para ver esto, considere la figura 15.7. Como muestra esta figura, si no se observa $Y$ (debido a la censura), todas esas observaciones ($= n_2$), denotadas por cruces, quedarán sobre el eje horizontal. Si se observa $Y$, las observaciones ($= n_1$) (señaladas con puntos) quedarán en el plano $X$-$Y$. La intuición indica que si estimamos una regresión basada sólo en las $n_1$ observaciones, los coeficientes resultantes del intercepto y de la pendiente estarán limitados a ser diferentes de los que obtendríamos si se tomaran en cuenta todas las $(n_1 + n_2)$ observaciones.

¿Cómo estimar entonces los modelos de regresión tobit (o censurados), como (15.11.1)? El mecanismo real implica al método de máxima verosimilitud, que, por su complejidad, escapa al alcance de este libro. Pero el lector puede obtener más información respecto del método MV en la bibliografía.[38]

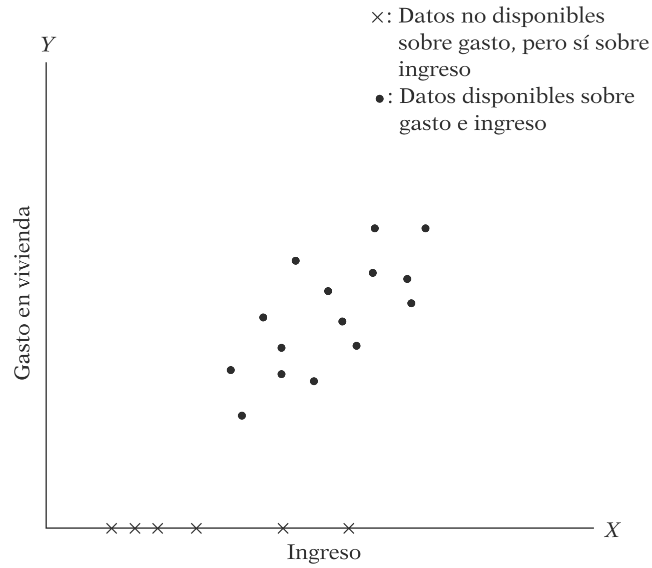
---

[36] Una muestra censurada debe diferenciarse de una **muestra truncada**, en la cual la información sobre las regresoras sólo está disponible si se observa la variable regresada. No analizaremos este tema aquí, pero el lector puede consultar William H. Greene, *Econometric Analysis*, Prentice Hall, 4a. ed., Englewood Cliffs, Nueva Jersey, cap. 19. Para un análisis intuitivo, véase Peter Kennedy, *A Guide to Econometrics*, The MIT Press, Cambridge, Massachusetts, 4a. ed., 1998, capítulo 16.

[37] El sesgo surge porque si sólo se consideran las $n_i$ observaciones y se omiten las demás, no hay garantía de que $E(u_i)$ será necesariamente igual a cero. Y sin $E(u_i) = 0$ no podemos garantizar que los estimadores de MCO serán insesgados. Este sesgo se ve fácilmente en el análisis del apéndice 3A, ecuaciones (4) y (5).

[38] Véase Greene, *op. cit.* Hay un análisis un poco menos técnico en Richard Breen, *Regression Models: Censores, Sampled Selected or Truncated Data*, Sage, Newbury Park, California, 1996.

**FIGURA 15.7**
Gráfico de la cantidad de dinero que gasta el consumidor en comprar una casa contra el ingreso.




James Heckman propuso un método alterno y más sencillo que el MV.[39] Consiste en un cálculo de dos pasos. En el primero estimamos la probabilidad de que un consumidor tenga una casa propia, con base en el modelo probit. En el siguiente paso estimamos el modelo (15.11.1) añadiéndole una variable (llamada **razón inversa de Mills** o **razón de riesgo**), la cual se deriva a partir de la estimación probit. Para conocer el mecanismo real, consulte el artículo de Hackman. El procedimiento Hackman proporciona estimaciones consistentes de los parámetros de (15.11.1), pero no tan eficientes como las estimaciones de MV. Como los programas estadísticos más modernos cuenta con rutinas de MV, serían preferibles en vez del proceso Hackman de dos pasos.

## Ilustración del modelo tobit: modelo de Ray Fair para las relaciones extramaritales[40]

En un interesante y novedoso artículo teórico, Ray Fair recopiló una muestra de 601 hombres y mujeres casados por primera vez y analizó sus respuestas a la pregunta respecto de relaciones extramaritales.[41] Las variables de este estudio se definen como sigue:

$Y =$ número de relaciones extramaritales durante el año anterior, 0, 1, 2, 3, 4-10 (codificado como 7)
$Z_1 =$ 0 para la mujer y 1 para el hombre
$Z_2 =$ edad
$Z_3 =$ número de años de matrimonio
$Z_4 =$ hijos: 0 si no los hay, 1 si los hay
$Z_5 =$ religiosidad en una escala de 1 a 5, 1 para los no religiosos
$Z_6 =$ escolaridad en años: básica $= 9$, licenciatura $= 12$, doctorado u otro grado $= 20$
$Z_7 =$ ocupación, escala "Hollingshead", de 1 a 7
$Z_8 =$ autovaloración del matrimonio, $1 =$ muy infeliz, $5 =$ muy feliz

[39] J.J. Heckman, "Simple Selection Bias as a Specification Error", *Econometrica*, vol. 47, pp. 153-161.

[40] Ray Fair, "A Theory of Extramarital Affaires", *Journal of Political Economy*, vol. 86, 1978, pp. 45-61. Para el artículo y los datos, consulte http://fairmodel.econ.yale.edu/rayfair/pdf/1978DAT.ZIP.

[41] En 1969, *Psychology Today* publicó un cuestionario con 101 preguntas sobre el sexo y pidió a sus lectores que enviaran por correo las respuestas. En el número correspondiente a julio de 1970 se analizaron los resultados de la encuesta con base en 2 000 respuestas que se recopilaron de forma electrónica. Ray Fair extrajo la muestra de 601 casos de estas respuestas.

**TABLA 15.18**
**Estimaciones por MCO y tobit de las relaciones extramaritales**

| Variable explicativa | Estimaciones MCO | Estimaciones tobit |
|---|---|---|
| Intercepto | 5.8720 (5.1622)* | 7.6084 (1.9479)† |
| $Z_1$ | 0.0540 (0.1799) | 0.9457 (0.8898) |
| $Z_2$ | −0.0509 (−2.2536) | −0.1926 (−2.3799) |
| $Z_3$ | 0.1694 (4.1109) | 0.5331 (3.6368) |
| $Z_4$ | −0.1426 (−0.4072) | 1.0191 (0.7965) |
| $Z_5$ | −0.4776 (−4.2747) | −1.6990 (−4.1906) |
| $Z_6$ | −0.0137 (−0.2143) | 0.0253 (0.1113) |
| $Z_7$ | 0.1049 (1.1803) | 0.2129 (0.6631) |
| $Z_8$ | −0.7118 (−5.9319) | −2.2732 (−5.4724) |
| $R^2$ | 0.1317 | 0.1515 |

* Las cifras entre paréntesis son los valores *t*.
† Las cifras entre paréntesis son los valores *Z* (normal estandarizados).
*Nota:* En total hay 601 observaciones, de las cuales 451 tienen valores cero para la variable dependiente (número de relaciones extramaritales), y 150, valores diferentes de cero.

De las 601 respuestas, 451 individuos no tuvieron relaciones extramaritales y 150 tuvieron una o más.

En términos de la figura 15.7, si graficamos el número de relaciones en el eje vertical y, por ejemplo, la escolaridad en el horizontal, habrá 451 observaciones a lo largo del eje horizontal. Por tanto, tenemos una muestra censurada, así que resulta apropiado un modelo tobit.

La tabla 15.18 proporciona las estimaciones del modelo anterior obtenidas mediante los procedimientos MCO (inapropiado) y MV (apropiado). Como se observa, el método de MCO incluye 451 individuos que no tuvieron relaciones y 150 que tuvieron una o más. El método MV toma esto en cuenta de manera explícita, pero los MCO no; he aquí la diferencia entre las dos estimaciones. Por razones ya vistas, debemos confiar en las estimaciones MV y no en las de MCO. Los coeficientes en ambos modelos pueden interpretarse como cualesquiera otros coeficientes de regresión. El coeficiente negativo de $Z_8$ (felicidad marital) significa que mientras más feliz se es, menor es la incidencia de relaciones extramaritales, hallazgo que quizá no sorprenda.

A propósito, observe que si nos interesa la probabilidad de las relaciones extramaritales y no su número, podemos utilizar el modelo probit, con $Y = 0$ para los individuos que no tuvieron relaciones de ese tipo y $Y = 1$ para los que sí las tuvieron, cuyos resultados se muestran en la tabla 15.19. Si ya saben elaborar modelos probit, los lectores deben ser capaces de interpretar los resultados probit de dicha tabla.

## 15.12 Creación de modelos para datos de cuenta: modelo de regresión de Poisson

Existen muchos fenómenos en los que la regresada es del **tipo de cuenta**, como el número de vacaciones tomadas por una familia en un año, el número de patentes otorgadas a una empresa en un año, el número de visitas a un dentista o a un doctor en un año, el número de visitas a un supermercado en una semana, el número de infracciones por estacionarse mal o conducir con exceso de velocidad en un año, el número de días en un hospital durante un periodo determinado, la cantidad de automóviles que pasan por una caseta en un intervalo de, por ejemplo, cinco minutos, etc. La variable en cada caso es discreta: toma sólo un número finito de valores. A veces los datos de cuenta se refieren a ocurrencias *raras* o *poco frecuentes*, como ser alcanzado por un rayo en el lapso de una semana, ganar más de dos veces la lotería en dos semanas o tener uno o más ataques al corazón en el transcurso de cuatro semanas. ¿Cómo elaboramos los modelos de estos fenómenos?

**TABLA 15.19**

Variable dependiente: YSTAR
Método: Logit binario MV
Muestra: 1-601
Observaciones incluidas: 601
Convergencia lograda después de 5 iteraciones

| Variable | Coeficiente | Error estándar | Estadístico $Z$ | Probabilidad |
|---|---|---|---|---|
| $C$ | 0.779402 | 0.512549 | 1.520638 | 0.1284 |
| $Z_1$ | 0.173457 | 0.137991 | 1.257015 | 0.2087 |
| $Z_2$ | -0.024584 | 0.010418 | -2.359844 | 0.0183 |
| $Z_3$ | 0.054343 | 0.018809 | 2.889278 | 0.0039 |
| $Z_4$ | 0.216644 | 0.165168 | 1.311657 | 0.1896 |
| $Z_5$ | -0.185468 | 0.051626 | -3.592551 | 0.0003 |
| $Z_6$ | 0.011262 | 0.029517 | 0.381556 | 0.7028 |
| $Z_7$ | 0.013669 | 0.041404 | 0.330129 | 0.7413 |
| $Z_8$ | -0.271791 | 0.053475 | -5.082608 | 0.0000 |

| | | | |
|---|---|---|---|
| Media de la variable dependiente | 0.249584 | Desviación estándar de la variable dependiente | 0.433133 |
| Error estándar de la regresión | 0.410279 | Criterio de información de Akaike | 1.045584 |
| Suma de cuadrados residual | 99.65088 | Criterio de Schwarz | 1.111453 |
| Log verosimilitud | -305.1980 | Criterio de Hannan-Quinn | 1.071224 |
| Log verosimilitud restr. | -337.6885 | Promedio log verosimilitud | -0.507817 |
| Estadístico RV (8 gl) | 64.98107 | $R$ cuadrada McFadden | 0.096215 |
| Probabilidad (estadístico RV) | 4.87E-11 | | |
| Obs. con Dep = 0 | 451 | Total de obs. | 601 |
| Obs. con Dep = 1 | 150 | | |

Del mismo modo que elegimos la distribución de Bernoulli para el modelo de decisiones del tipo sí/no en el modelo lineal de probabilidad, la distribución de probabilidades específicamente adecuada para los datos de cuenta es la distribución de probabilidades de **Poisson**. La fdp de la distribución de Poisson está dada por:[42]

$$f(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!} \qquad Y = 0, 1, 2, \ldots \tag{15.12.1}$$

donde $f(Y)$ denota la probabilidad de que la variable $Y$ tome valores enteros no negativos, y donde $Y!$ (se lee como $Y$ factorial) significa $Y! = Y \times (Y-1) \times (Y-2) \times 2 \times 1$. Se puede demostrar que

$$E(Y) = \mu \tag{15.12.2}$$

$$\operatorname{var}(Y) = \mu \tag{15.12.3}$$

Observe una característica importante de la distribución de Poisson: *su varianza es la misma que el valor de su media.*

El modelo de regresión de Poisson se expresa como:

$$Y_i = E(Y_i) + u_i = \mu_i + u_i \tag{15.12.4}$$

[42] Consulte cualquier libro usual de estadística para los detalles de esta distribución.

donde las $Y$ están independientemente distribuidas como variables aleatorias Poisson, con una media $\mu_i$ para cada individuo expresada como

$$\mu_i = E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3_i} + \cdots + \beta_k X_{ki} \tag{15.12.5}$$

donde las $X$ son algunas variables que afectarían el valor de la media. Por ejemplo, si la variable de cuenta es el número de visitas al Museo Metropolitano de Nueva York en un año determinado, esta cifra dependerá de variables como el ingreso del consumidor, el precio de entrada, la distancia al museo y las tarifas de estacionamiento.

Para propósitos de cálculo, expresamos el modelo como

$$Y_i = \frac{\mu^Y e^{-\mu}}{Y!} + u_i \tag{15.12.6}$$

en donde $\mu$ está sustituida por (15.12.5). Como se aprecia fácilmente, el modelo de regresión resultante será no lineal en los parámetros, por lo que requiere una estimación de regresión no lineal, que analizamos en el capítulo anterior. Consideremos un ejemplo concreto para ver cómo funciona esto.

**EJEMPLO 15.8**
*Un ejemplo ilustrativo: estudio geriátrico sobre la frecuencia de caídas*

Estos datos se recopilaron por Neter *et al.*[43] Los datos se refieren a 100 individuos de 65 años de edad o mayores. El objetivo del estudio fue registrar el número de caídas ($= Y$) sufridas por estos individuos según el sexo ($X_2 = 0$ si es mujer y 1 si es hombre), índice de equilibrio ($X_3$) e índice de fortaleza ($X_4$). Mientras mayor sea el índice de equilibrio, más estable será el sujeto; y mientras mayor sea el índice de fortaleza, más fuerte será el individuo. Para averiguar si la escolaridad, o la escolaridad más los ejercicios aeróbicos, influyen en el número de caídas, los autores introdujeron la variable adicional ($X_1$), llamada variable de *intervención*, la cual $X_1 = 0$ si sólo interviene la escolaridad, y $X_1 = 1$ si se trata de la escolaridad más los ejercicios aeróbicos. Los sujetos se asignaron de manera aleatoria a los dos métodos de intervención.

Con *EViews 6* obtuvimos los resultados de la tabla 15.20.

**TABLA 15.20**

Variable dependiente: *Y*
Muestra: 1-100
Convergencia lograda después de 7 iteraciones
*Y*=EXP(*C*(0)+*C*(1)**X*1+*C*(2)**X*2+*C*(3)**X*3+*C*(4)**X*4)

| | Coeficiente | Error estándar | Estadístico *t* | Probabilidad |
|---|---|---|---|---|
| *C*(0) | 0.37020 | 0.3459 | 1.0701 | 0.2873 |
| *C*(1) | -1.10036 | 0.1705 | -6.4525 | 0.0000 |
| *C*(2) | -0.02194 | 0.1105 | -0.1985 | 0.8430 |
| *C*(3) | 0.01066 | 0.0027 | 3.9483 | 0.0001 |
| *C*(4) | 0.00927 | 0.00414 | 2.2380 | 0.0275 |

$R^2$ = 0.4857 $R^2$ ajustada = 0.4640
Log verosimilitud = -197.2096 Estadístico de Durbin-Watson = 1.7358

*Nota: EXP*( ) significa *e* (la base del logaritmo natural) elevado a la expresión entre paréntesis.

[43] John Meter, Michael H. Kutner, Christopher J. Nachtsheim y William Wasserman, *Applied Regression Models*, Irwin, 3a. ed., Chicago, 1996. Los datos provienen del disco de datos incluido en el libro y se refieren al ejercicio 14.28.

**EJEMPLO 15.8** (*continuación*)

**Interpretación de los resultados.** Tenga en cuenta que lo que obtuvimos en la tabla 15.20 es el valor medio estimado para el *i*-ésimo individuo, $\hat{\mu}_i$; es decir, lo que estimamos es:

$$\hat{\mu}_i = e^{0.3702-1.100366X_{1i}-0.02194X_{2i}+0.0106X_{3i}+0.00927X_{4i}} \tag{15.12.7}$$

Para encontrar el valor medio real del *i*-ésimo sujeto necesitamos colocar los valores de las distintas variables $X$ de ese sujeto. Por ejemplo, el sujeto 99 tuvo estos valores: $Y = 4$, $X_1 = 0$, $X_2 = 1$, $X_3 = 50$ y $X_4 = 56$. Al colocar dichos valores en (15.12.7), obtenemos $\hat{\mu}_{99} = 3.3538$ como valor medio estimado del sujeto 99. El valor real $Y$ de este individuo fue 4.

Ahora bien, si deseamos saber la probabilidad de que un sujeto similar al 99 sufra menos de cinco caídas al año, se obtiene mediante:

$$\begin{aligned} P(Y < 5) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) \\ &= \frac{(3.3538)^0 e^{-3.3538}}{0!} + \frac{(3.3538)^1 e^{-3.3538}}{1!} + \frac{(3.3538)^2 e^{-3.3538}}{2!} \\ &\quad + \frac{(3.3538)^3 e^{-3.3538}}{3!} + \frac{(3.3538)^4 e^{-3.3538}}{4!} \\ &= 0.7491 \end{aligned}$$

Asimismo, calculamos el efecto marginal o parcial que una regresora tiene en el valor medio de $Y$ de la siguiente manera. En términos de este ejemplo, suponga que deseamos averiguar el efecto de un incremento de una unidad en el índice de fortaleza ($X_4$) sobre la media $Y$. Como

$$\mu = e^{C_0+C_1X_{1i}+C_2X_{2i}+C_3X_{3i}+C_4X_{4i}} \tag{15.12.8}$$

deseamos conocer $\partial\mu/\partial X_4$. Con la regla de la cadena del cálculo, demostramos fácilmente que lo anterior es igual a

$$\frac{\partial\mu}{\partial X_4} = C_4 e^{C_0+C_1X_{1i}+C_2X_{2i}+C_3X_{3i}+C_4X_{4i}} = C_4\mu \tag{15.12.9}$$

Es decir, la tasa de cambio del valor medio respecto de la regresora es igual al coeficiente de esa regresora multiplicado por el valor medio. Por supuesto, el valor medio $\mu$ depende de los valores tomados por todas las regresoras en el modelo. Esto es similar a los modelos probit y logit ya analizados, en los que la contribución marginal de una variable también dependía de los valores tomados por todas las variables del modelo.

De regreso a la importancia estadística de los coeficientes individuales, observamos que el intercepto y la variable $X_2$ son estadísticamente significativas en lo individual. Pero note que los errores estándar dados en la tabla son asintóticos y, por tanto, los valores $t$ deben interpretarse de manera asintótica. Como ya mencionamos, por lo general los resultados de todos los procedimientos iterativos de estimación no lineales sólo tienen validez en muestras grandes.

Para concluir el análisis del modelo de regresión de Poisson, vale la pena mencionar que el modelo hace supuestos restrictivos, como el que la media y la varianza del proceso de Poisson son iguales y que la probabilidad de una ocurrencia es constante en cualquier punto en el tiempo.

## 15.13 Otros temas de los modelos de regresión de respuesta cualitativa

Como expresamos al principio, el tema de los modelos de regresión con respuesta cualitativa es amplio. Lo que presentamos en este capítulo son algunos modelos básicos de este tema. Para quienes deseen adentrarse más en esta área, a continuación estudiaremos muy brevemente otros modelos. No profundizaremos en ellos, pues están fuera del alcance de esta obra.

## Modelos ordinales logit y probit

En los modelos bivariados logit y probit, el interés residía en construir un modelo para una variable de respuesta del tipo sí o no. Pero a menudo la variable de respuesta, o regresada, puede tener más de dos resultados, y con mucha frecuencia son **ordinales** por naturaleza; es decir, no pueden expresarse en una escala de intervalo. Suele suceder que en las investigaciones del tipo de encuesta las respuestas se den en una escala de tipo Likert, por ejemplo, "totalmente de acuerdo", "algo de acuerdo", o "totalmente en desacuerdo". O las respuestas de una encuesta sobre educación quizá sean "menor a la educación media superior", "educación media superior", "licenciatura", "posgrado". Muy a menudo tales respuestas se codifican como 0 (menor a la educación media superior), 1 (educación media superior), 2 (licenciatura) y 3 (posgrado). Éstas son escalas ordinales, pues hay un orden claro entre las categorías, pero no podemos decir que 2 (licenciatura) es dos veces 1 (educación media superior), o que 3 (posgrado) es tres veces 1 (educación media superior).

Para estudiar estos fenómenos extendemos los modelos bivariados logit y probit a fin de que tomen en cuenta múltiples categorías ordenadas. La aritmética tiene mucho que ver, pues se precisan distribuciones de probabilidades logística y normal en múltiples etapas para las diversas categorías ordenadas. Respecto de las matemáticas necesarias y algunas de sus aplicaciones, el lector puede consultar los textos de Greene y Maddala ya citados. En un nivel comparativamente intuitivo, puede consultar la monografía de Liao.[44] Los programas de software como LIMDEP, *EViews,* STATA y SHAZAM tienen rutinas para estimar los modelos logit y probit ordenados.

## Modelos multinomiales logit y probit

En los modelos probit y logit ordenados, las variables de respuesta tienen más de dos categorías ordenadas o jerarquizadas. Pero existen situaciones en las que la regresada no está ordenada. Por ejemplo, considere la elección del transporte para ir al trabajo. Las elecciones son bicicleta, motocicleta, automóvil, autobús o tren. Aunque son respuestas categóricas, no existe jerarquía ni orden; en esencia, tienen una naturaleza nominal. Otro ejemplo: las clasificaciones laborales, como mano de obra no especializada, semiespecializada y muy especializada. De nuevo, no hay ningún orden. De manera semejante, las elecciones ocupacionales, como autoempleado, empleado de una empresa privada, empleado de una oficina gubernamental local y empleado de una oficina gubernamental federal, son de carácter esencialmente nominal.

Las técnicas de los modelos probit y logit multinomiales sirven para estudiar esas categorías nominales. Una vez más, las matemáticas intervienen en cierta medida. Las referencias ya citadas ofrecen los elementos esenciales de esas técnicas. Asimismo, el software estadístico ya mencionado es útil para aplicar tales modelos si así se requiere en casos específicos.

## Modelos de duración

Considere preguntas como las siguientes: 1) ¿qué determina la duración de los intervalos de desempleo?, 2) ¿qué determina la vida de un foco?, 3) ¿qué factores determinan la duración de una huelga?, 4) ¿qué determina el tiempo de sobrevivencia de un paciente VIH positivo?

Materias como las anteriores son el tema de los modelos de duración, popularmente conocidos como **análisis de supervivencia** o **análisis de datos del tiempo a un suceso**. En cada ejemplo citado, la variable clave es la longitud del tiempo o la longitud del intervalo, cuyo modelo es una variable aleatoria. Una vez más, las matemáticas implican las FDP y las FDA de distribuciones de probabilidades apropiadas. Aunque los detalles técnicos pueden resultar tediosos, hay libros accesibles sobre la materia.[45] El software estadístico como STATA y LIMDEP estima con faci-

---

[44] Tim Futing Liao, *op. cit*.

[45] Véase, por ejemplo, David W. Hosmer, Jr., y Stanley Lemeshow, *Applied Survival Analysis*, John Wiley & Sons, Nueva York, 1999.

lidad tales modelos de duración. Estos paquetes cuentan con ejemplos resueltos para ayudar al investigador con tales modelos.

## Resumen y conclusiones

1. Los modelos de regresión con respuesta cualitativa se refieren a modelos en los que la variable de respuesta, o regresada, no es cuantitativa ni en escala de intervalo.
2. El modelo de regresión con respuesta cualitativa más sencillo posible es el modelo binario en el que la regresada es del tipo sí/no o presencia/ausencia.
3. El modelo de regresión binario más sencillo posible es el modelo lineal de probabilidad (MLP), en el que se hace la regresión sobre la variable de respuesta binaria con la metodología de MCO estándar. En este caso, la simplicidad quizá no sea una virtud, pues el MLP experimenta diversos problemas de estimación. Aunque se superen algunos de dichos problemas de estimación, la debilidad fundamental del MLP es que supone que la probabilidad de que algo suceda se incrementa de manera lineal en función del nivel de la regresora; este supuesto tan restrictivo se evita con los modelos probit y logit.
4. En el modelo logit, la variable dependiente es el logaritmo de la razón de probabilidades, la cual es una función lineal de las regresoras. La función de probabilidades del modelo logit es la distribución logística. Si se contara con los datos de manera agrupada, se utilizarían MCO para calcular los parámetros del modelo logit, siempre y cuando se tome en cuenta de manera explícita la naturaleza heteroscedástica del término de error. Si se dispone de los datos en el nivel individual o micro, se requerirían los procedimientos de estimación no lineales en los parámetros.
5. Si elegimos la distribución normal como la distribución de probabilidades apropiada, se emplea el modelo probit, aunque es matemáticamente más difícil porque requiere integrales. Pero para propósitos prácticos, los resultados de los modelos logit y probit son similares. En la práctica, la elección depende de la facilidad de cálculo, lo cual no representa un problema grave en vista del complejo software estadístico que hay ahora.
6. Si la variable de respuesta es del tipo de cuenta, el modelo más frecuente en el trabajo aplicado es el de regresión de Poisson, que se basa en la distribución de probabilidades de Poisson.
7. Un modelo estrechamente relacionado con el modelo probit es el tobit, también conocido como modelo de regresión censurado. En dicho modelo, la variable de respuesta se observa sólo si se cumple(n) cierta(s) condición(es). Así, la pregunta de qué cantidad se gasta en un automóvil resulta significativa sólo si, para empezar, se decide adquirir un automóvil. Sin embargo, Maddala observa que el modelo tobit es "aplicable sólo en esos casos en donde la variable latente [es decir, la variable básica que subyace en un fenómeno] puede, en principio, adoptar valores negativos, y los valores nulos observados son una consecuencia de la censura y la no observabilidad".[46]
8. Existen varias extensiones del modelo de regresión con respuesta binaria, como los modelos probit y logit ordenados, así como los probit y logit nominales. La filosofía de estos modelos es la misma que la de los modelos logit y probit más sencillos, a pesar de que las matemáticas se complican un poco.
9. Por último, mencionamos brevemente los llamados modelos de duración, en los que la duración de un fenómeno, como el desempleo o la enfermedad, depende de diversos factores. En tales modelos, la longitud o el intervalo de duración se convierten en una variable de interés para la investigación.

[46] G.S. Maddala, *Introduction to Econometrics*, 2a. ed., Macmillan, Nueva York, 1992, p. 342.

# EJERCICIOS

## Preguntas

15.1. Consulte la información de la tabla 15.2. Si $\hat{Y}_i$ es negativa, suponga que es igual a 0.01, y si es mayor que 1, suponga que es igual a 0.99. Recalcule las ponderaciones $w_i$ y estime el MLP mediante MCP. Compare los resultados con los dados en (15.2.11) y coméntelos.

15.2. Para la información sobre propiedad de vivienda de la tabla 15.1, las estimaciones de máxima verosimilitud del modelo logit son las siguientes:

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -493(54 + \quad 32.96 \text{ ingreso}$$
$$t = \quad (-0.000008)(0.000008)$$

Comente estos resultados, teniendo en mente que todos los valores de ingreso por encima de 16 (1 000 dólares) corresponden a $Y = 1$ y todos los valores de ingreso por debajo de 16 corresponden a $Y = 0$. *A priori*, ¿qué esperaría en tal situación?

15.3. Al estudiar la compra de bienes imperecederos $Y$ ($Y = 1$ si hubo compra, $Y = 0$ si no la hubo) como función de diversas variables para un total de 762 familias, Janet A. Fisher* obtuvo los siguientes resultados del MLP:

| Variable explicativa | Coeficiente | Error estándar |
|---|---|---|
| Constante | 0.1411 | — |
| Ingreso disponible 1957, $X_1$ | 0.0251 | 0.0118 |
| (Ingreso disponible $= X_1)^2$, $X_2$ | −0.0004 | 0.0004 |
| Cuentas corrientes, $X_3$ | −0.0051 | 0.0108 |
| Cuentas de ahorro, $X_4$ | 0.0013 | 0.0047 |
| Bonos de ahorro de Estados Unidos, $X_5$ | −0.0079 | 0.0067 |
| Condición de vivienda: alquilada, $X_6$ | −0.0469 | 0.0937 |
| Condición de vivienda: propia, $X_7$ | 0.0136 | 0.0712 |
| Alquiler mensual, $X_8$ | −0.7540 | 1.0983 |
| Pagos de hipoteca mensual, $X_9$ | −0.9809 | 0.5162 |
| Deuda personal pagadera en una sola exhibición, $X_{10}$ | −0.0367 | 0.0326 |
| Edad, $X_{11}$ | 0.0046 | 0.0084 |
| Edad elevada al cuadrado, $X_{12}$ | −0.0001 | 0.0001 |
| Estado civil, $X_{13}$ (1 = casado) | 0.1760 | 0.0501 |
| Número de hijos, $X_{14}$ | 0.0398 | 0.0358 |
| (Número de hijos $= X_{14})^2$, $X_{15}$ | −0.0036 | 0.0072 |
| Planes de compra, $X_{16}$ (1 = planeado; 0 en otro caso) | 0.1760 | 0.0384 |
| $R^2 = 0.1336$ | | |

*Notas:* Todas las variables financieras están expresadas en miles de dólares.
Condición de vivienda: alquilada (1 si es alquilada; 0 en otro caso).
Condición de vivienda: propia (1 si es propia; 0 en otro caso).

Fuente: Janet A. Fisher, "An Analysis of Consumer Good Expenditure", *The Review of Economics and Statistics*, vol. 64, núm. 1, tabla 1, 1962, p. 67.

*a*) Comente en general sobre el ajuste de la ecuación.

*b*) ¿Cómo interpreta el coeficiente de −0.0051 asociado a la variable de cuentas corrientes? ¿Cómo explica el signo negativo de esta variable?

*c*) ¿Cuál es el razonamiento de la introducción de las variables edad elevada al cuadrado y número de hijos elevado al cuadrado? ¿Por qué hay signo negativo en ambos casos?

* "An Analysis of Consumer Good Expenditure", *The Review of Economics and Statistics*, vol. 64, núm. 1, 1962, pp. 64-71.

*d*) Si tiene valores de cero para todas las variables excepto para la variable ingreso, encuentre la probabilidad condicional de una familia, cuyo ingreso es \$20 000, de comprar un bien imperecedero.

*e*) Estime la probabilidad condicional de poseer uno o más bienes imperecederos si $X_1 = \$15\,000$, $X_3 = \$3\,000$, $X_4 = \$5\,000$, $X_6 = 0$, $X_7 = 1$, $X_8 = \$500$, $X_9 = \$300$, $X_{10} = 0$, $X_{11} = 35$, $X_{13} = 1$, $X_{14} = 2$, $X_{16} = 0$.

15.4. El valor $R^2$ en la regresión de la participación de la fuerza laboral en la tabla 15.3 es 0.175, relativamente bajo. ¿Puede probar la significancia estadística para este valor? ¿Qué prueba utiliza y por qué? Comente en general sobre el valor del $R^2$ en tales modelos.

15.5. Estime las probabilidades de tener casa propia en los diversos niveles de ingreso en los que se basa la regresión (15.7.1). Grafique las probabilidades contra el ingreso y comente sobre la relación resultante.

*15.6. En la regresión probit de la tabla 15.11 muestre que el intercepto es igual a $-\mu_x/\sigma_x$ y la pendiente es igual a $1/\sigma_x$, donde $\mu_x$ y $\sigma_x$ son la media y la desviación estándar de *X*.

15.7. Con base en datos de 54 áreas estadísticas metropolitanas estándar (AEME), Demaris estimó el siguiente modelo logit para explicar tasas altas de asesinatos *versus* tasas bajas de asesinatos:**

$$\begin{aligned} \ln \hat{O}_i &= 1.1387 + 0.0014P_i + 0.0561C_i - 0.4050R_i \\ \text{ee} &= \qquad\quad (0.0009) \quad (0.0227) \quad (0.1568) \end{aligned}$$

donde $O =$ posibilidades en favor de una tasa alta de asesinatos, $P =$ tamaño de la población en 1980 en miles, $C =$ tasa de crecimiento de la población entre 1970 y 1980, $R =$ cociente de lectura, y ee, errores estándar asintóticos.

*a*) ¿Cómo interpretaría los diversos coeficientes?

*b*) ¿Qué coeficientes son en lo individual estadísticamente significativos?

*c*) ¿Cuál es el efecto de un incremento unitario en el cociente de lectura en las posibilidades en favor de una tasa más alta de asesinatos?

*d*) ¿Cuál es el efecto de un incremento de un punto porcentual en la tasa de crecimiento poblacional sobre las posibilidades en favor de una tasa más alta de asesinatos?

15.8. Compare y comente sobre las regresiones de MCO y MCP (15.7.3) y (15.7.1).

## Ejercicios empíricos

15.9. De la encuesta sobre presupuesto familiar de 1980 levantada por la Oficina Central Holandesa de Estadísticas, J. S. Cramer obtuvo el siguiente modelo logit con base en una muestra de 2 820 familias. (Los resultados se basan en el método de máxima verosimilitud y se dan después de la tercera iteración.)[†] El propósito del modelo logit fue determinar la adquisición de un automóvil como una función del (logaritmo del) ingreso. La adquisición de automóvil fue una variable binaria: $Y = 1$ si una familia tenía un automóvil, $Y = 0$ en otro caso.

$$\begin{aligned} \hat{L}_i &= -2.77231 + 0.347582 \ln \text{Ingreso} \\ t &= (-3.35) \qquad (4.05) \\ \chi^2(1\text{ gl}) &= 16.681 \text{ (valor } p = 0.0000) \end{aligned}$$

donde $\hat{L}_i =$ el logit estimado y en donde ln Ingreso es el logaritmo del ingreso. La distribución $\chi^2$ mide la bondad del ajuste del modelo.

* Opcional.

** Demaris, *op. cit.*, p. 46.

[†] J.S. Cramer, *An Introduction to the Logit Model for Economist*, 2a. ed., publicada y distribuida por Timberlake Consultants, 2001, p. 33. Estos resultados se reproducen del software estadístico PC-GIVE 10, publicado por Timberlake Consultants, p. 51.

*a*) Interprete el modelo logit estimado.

*b*) Del modelo logit estimado, ¿cómo obtendría la expresión para la probabilidad de adquirir un automóvil?

*c*) ¿Cuál es la probabilidad de que una familia con un ingreso de $20 000 posea un automóvil?, ¿y para un nivel de ingreso de $25 000? ¿Cuál es la tasa de cambio de la probabilidad en un nivel de ingreso de $20 000?

*d*) Comente sobre la significancia estadística del modelo logit estimado.

15.10. Establezca la ecuación (15.2.8).

15.11. En un importante estudio sobre las tasas de graduación de todos los alumnos inscritos en educación media superior y de los inscritos afroestadounidenses únicamente, Bowen y Bok obtuvieron los resultados de la tabla 15.21, con base en el modelo logit.*

**TABLA 15.21** **Modelo de regresión logística para predecir tasas de graduación, Generación entrante 1989**

| | Todos los inscritos | | | Sólo afroestadounidenses | | |
|---|---|---|---|---|---|---|
| **Variable** | **Estimación del parámetro** | **Error estándar** | **Razón de probabilidad** | **Estimación del parámetro** | **Error estándar** | **Razón de probabilidad** |
| Intercepto | 0.957 | 0.052 | — | 0.455 | 0.112 | — |
| Mujeres | **0.280** | 0.031 | 1.323 | **0.265** | 0.101 | 1.303 |
| Negros | **–0.513** | 0.056 | 0.599 | | | |
| Hispanos | **–0.350** | 0.080 | 0.705 | | | |
| Asiáticos | **0.122** | 0.055 | 1.130 | | | |
| Otra raza | **–0.330** | 0.104 | 0.719 | | | |
| SAT > 1 299 | **0.331** | 0.059 | 1.393 | 0.128 | 0.248 | 1.137 |
| SAT 1 200-1 299 | **0.253** | 0.055 | 1.288 | 0.232 | 0.179 | 1.261 |
| SAT 1 100-1 199 | **0.350** | 0.053 | 1.420 | 0.308 | 0.149 | 1.361 |
| SAT 1 000-1 099 | **0.192** | 0.054 | 1.211 | 0.141 | 0.136 | 1.151 |
| SAT no disponible | **–0.330** | 0.127 | 0.719 | 0.048 | 0.349 | 1.050 |
| 10% superior de la generación de escuela media superior | **0.342** | 0.036 | 1.407 | **0.315** | 0.117 | 1.370 |
| Rango de la generación de escuela media superior no disponible | –0.065 | 0.046 | 0.937 | –0.065 | 0.148 | 0.937 |
| Nivel socioeconómico alto (NSE) | **0.283** | 0.036 | 1.327 | **0.557** | 0.175 | 1.746 |
| NSE bajo | **–0.385** | 0.079 | 0.680 | **–0.305** | 0.143 | 0.737 |
| NSE no disponible | **0.110** | 0.050 | 1.116 | 0.031 | 0.172 | 1.031 |
| SEL-1 | **1.092** | 0.058 | 2.979 | **0.712** | 0.161 | 2.038 |
| SEL-2 | **0.193** | 0.036 | 1.212 | **0.280** | 0.119 | 1.323 |
| Universidad para mujeres | **–0.299** | 0.069 | 0.742 | 0.158 | 0.269 | 1.171 |
| Número de observaciones | 32 524 | | | 2 354 | | |
| –2 log verosimilitud | | | | | | |
| Restringido | 31 553 | | | 2 667 | | |
| No restringido | 30 160 | | | 2 569 | | |
| Ji cuadrada | 1 393 con 18 gl | | | 98 con 14 gl | | |

*Notas:* Los coeficientes en negritas son significativos en el nivel de 0.05; los otros coeficientes no lo son. Las categorías omitidas en el modelo son blancos, hombres, SAT < 1 000, 90% inferior de la generación de escuela media superior, NSE medio, SEL-3, institución mixta. Las tasas de graduación se refieren a las tasas de graduación de 6 años de educación primaria, según se define en las notas de la tabla D.3.1 del apéndice. Las categorías de selectividad institucional se definen en las notas de la tabla D.3.1 del apéndice. Véase el **apéndice B** para la definición de nivel socioeconómico (NSE).

SEL-1 = instituciones con una media de calificaciones combinadas SAT de 1 300 o superior.

SEL-2 = instituciones con una media de calificaciones combinadas SAT entre 1 150 y 1 299.

SEL-1 = instituciones con una media de calificaciones combinadas SAT inferior a 1 150.

Fuente: Bowen y Bok, *op. cit*., p. 381.

* William G. Bowen y Derek Bok, *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*, Princeton University Press, Princeton, Nueva Jersey, 1998, p. 381.

*a*) ¿Qué conclusión general se deduce sobre las tasas de graduación de todos los inscritos y de únicamente los inscritos afroestadounidenses?

*b*) La *razón de probabilidad* es la razón de dos probabilidades. Compare dos grupos de todos los inscritos, uno con una calificación SAT mayor que 1 299 y otro con una calificación SAT menor que 1 000 (la categoría base). La razón de probabilidad de 1.393 significa que las posibilidades en favor de que los inscritos en la primera categoría terminen sus estudios universitarios son 39% superiores a las de los que están en la otra categoría. ¿Las diversas razones de probabilidad que se muestran en la tabla están de acuerdo con las expectativas *a priori*?

*c*) ¿Qué puede decir respecto de la significancia estadística de los parámetros estimados? ¿Y respecto de la importancia global del modelo estimado?

15.12. En el modelo probit (tabla 15.11), la perturbación $u_i$ tiene esta varianza:

$$\sigma_u^2 = \frac{P_i(1 - P_i)}{N_i f_i^2}$$

donde $f_i$ es la función de densidad normal estándar evaluada en $F^{-1}(P_i)$.

*a*) Con la varianza anterior de $u_i$, ¿cómo transformaría el modelo de la tabla 15.10 para que el término de error resultante sea homoscedástico?

*b*) Con la información de la tabla 15.10, muestre la información transformada.

*c*) Estime el modelo probit con base en los datos transformados y compare los resultados con los basados en los datos originales.

15.13. Como $R^2$ en calidad de medida de bondad de ajuste no es particularmente apropiada para los modelos de variable dependiente dicótoma, una alternativa que se propone es la prueba $\chi^2$ descrita en seguida:

$$\chi^2 = \sum_{i=1}^{G} \frac{N_i(\hat{P}_i - P_i^*)^2}{P_i^*(1 - P_i^*)}$$

donde $N_i$ = número de observaciones en la *i*-ésima celda
$\hat{P}_i$ = probabilidad real de ocurrencia del suceso ($= n_i/N_i$)
$P_i^*$ = probabilidad estimada
$G$ = número de celdas (es decir, el número de niveles en el cual se mide $X_i$, por ejemplo, 10 en la tabla 15.4)

Puede demostrarse que, para muestras grandes, $\chi^2$ está distribuida de acuerdo con la distribución $\chi^2$ con $(G - k)$ gl, donde $k$ es el número de parámetros en el modelo por estimar $(k < G)$.

Aplique la prueba anterior $\chi^2$ a la regresión (15.7.1), comente la bondad de ajuste resultante y compare con el valor reportado de la $R^2$.

15.14. La tabla 15.22 presenta datos sobre los resultados de dispersar rotenona en diferentes concentraciones sobre poblaciones de aproximadamente cincuenta individuos del pulgón del crisantemo. Elabore un modelo adecuado para expresar la probabilidad de mortalidad como una función del logaritmo de $X$, el logaritmo de la dosificación, y comente los resultados. También calcule la prueba $\chi^2$ del ajuste estudiado en el ejercicio 15.13.

15.15. Trece aspirantes a un programa de posgrado obtuvieron las calificaciones cuantitativa y verbal en el examen GRE que se presentan en la tabla 15.23. Seis estudiantes fueron admitidos en el programa.

*a*) Con el modelo MLP, prediga la probabilidad de admisión al programa con base en las calificaciones cuantitativas y verbales del GRE.

*b*) ¿Es un modelo satisfactorio? De no ser así, ¿qué alternativa(s) ofrece?

**TABLA 15.22**
**Estudio de toxicidad de rotenona en poblaciones de pulgón del crisantemo**

Fuente: D. J. Fennet, *Probit Analysis*, Cambridge University Press, Londres, 1964.

| Concentración, miligramos por litro | | Total | Mortalidad, | |
|---|---|---|---|---|
| $X$ | log ($X$) | $N_i$ | $n_i$ | $\hat{P}_i = n_i/N_i$ |
| 2.6 | 0.4150 | 50 | 6 | 0.120 |
| 3.8 | 0.5797 | 48 | 16 | 0.333 |
| 5.1 | 0.7076 | 46 | 24 | 0.522 |
| 7.7 | 0.8865 | 49 | 42 | 0.857 |
| 10.2 | 1.0086 | 50 | 44 | 0.880 |

**TABLA 15.23**
**Calificaciones del examen GRE**

Fuente: Donald F. Morrison, *Applied Linear Statistical Methods*, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1983, p. 279 (adaptada).

| | Calificaciones de examen de aptitud GRE | | Admitido al |
|---|---|---|---|
| Núm. de estudiante | Cuantitativo, $Q$ | Verbal, $V$ | programa de posgrado (Sí = 1, No = 0) |
| 1 | 760 | 550 | 1 |
| 2 | 600 | 350 | 0 |
| 3 | 720 | 320 | 0 |
| 4 | 710 | 630 | 1 |
| 5 | 530 | 430 | 0 |
| 6 | 650 | 570 | 0 |
| 7 | 800 | 500 | 1 |
| 8 | 650 | 680 | 1 |
| 9 | 520 | 660 | 0 |
| 10 | 800 | 250 | 0 |
| 11 | 670 | 480 | 0 |
| 12 | 670 | 520 | 1 |
| 13 | 780 | 710 | 1 |

15.16. Para estudiar la eficacia de un cupón de descuento sobre el precio de un paquete de seis unidades de un refresco, Douglas Montgomery y Elizabeth Peck recopilaron los datos que aparecen en la tabla 15.24. Se asignó aleatoriamente una muestra de 5 500 consumidores a las once categorías de descuento que aparecen en la tabla, 500 por categoría. La variable de respuesta es si los consumidores canjearon el cupón en un lapso de un mes, o si no lo hicieron.

*a*) Vea si el modelo logit se ajusta a la información; considere la tasa de canje como variable dependiente y el precio de descuento como variable explicativa.

*b*) Vea si el modelo probit opera tan bien como el modelo logit.

**TABLA 15.24**
**Precio del refresco con el cupón de descuento**

Fuente: Douglas C. Montgomery y Elizabeth A. Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, Nueva York, 1982, p. 243 (la notación cambió).

| Precio con descuento $X$, ¢ | Tamaño de la muestra $N_i$ | Número de cupones canjeados $n_i$ |
|---|---|---|
| 5 | 500 | 100 |
| 7 | 500 | 122 |
| 9 | 500 | 147 |
| 11 | 500 | 176 |
| 13 | 500 | 211 |
| 15 | 500 | 244 |
| 17 | 500 | 277 |
| 19 | 500 | 310 |
| 21 | 500 | 343 |
| 23 | 500 | 372 |
| 25 | 500 | 391 |

*c*) ¿Cuál es la tasa de canje proyectada si la tasa de descuento de precio fue de 17 centavos?

*d*) Estime el precio de descuento por el cual se canjeará 70% de los cupones.

15.17. Para averiguar quién tiene una cuenta bancaria (corriente, de ahorro, etc.,) y quién no la tiene, John Caskey y Andrew Peterson estimaron un modelo probit para los años 1977 y 1989, con información sobre las familias en Estados Unidos. Los resultados están en la tabla 15.25. Los valores de los coeficientes de pendiente en la tabla miden el efecto de un cambio unitario en una regresora sobre la probabilidad de que una familia tenga una cuenta bancaria, y con el cálculo estos efectos marginales en los valores medios de las regresoras del modelo.

*a*) Para 1977, ¿cuál es el efecto del estado civil sobre la propiedad de una cuenta bancaria?, ¿y para 1989? ¿Tienen sentido económico estos resultados?

*b*) ¿Por qué es negativo el coeficiente asociado a la variable minoría para 1977 al igual que para 1989?

*c*) ¿Cómo explica el signo negativo de la variable número de hijos?

*d*) ¿Qué indica el estadístico ji cuadrada dado en la tabla? (*Sugerencia:* Consulte el ejercicio 15.13.)

**TABLA 15.25** **Regresiones probit donde la variable dependiente es la posesión de una cuenta bancaria**

| | Datos de 1977 | | Datos de 1989 | |
|---|---|---|---|---|
| | **Coeficientes** | **Pendiente implícita** | **Coeficientes** | **Pendiente implícita** |
| Constante | –1.06 | | –2.20 | |
| | (3.3)* | | (6.8)* | |
| Ingreso (miles de dólares de 1991) | 0.030 | 0.002 | 0.025 | 0.002 |
| | (6.9) | | (6.8) | |
| Casado | 0.127 | 0.008 | 0.235 | 0.023 |
| | (0.8) | | (1.7) | |
| Número de hijos | –0.131 | –0.009 | –0.084 | –0.008 |
| | (3.6) | | (2.0) | |
| Edad del jefe de familia (JF) | 0.006 | 0.0004 | 0.021 | 0.002 |
| | (1.7) | | (6.3) | |
| Educación de JF | 0.121 | 0.008 | 0.128 | 0.012 |
| | (7.4) | | (7.7) | |
| JF hombre | –0.078 | –0.005 | –0.144 | –0.011 |
| | (0.5) | | (0.9) | |
| Minoría | –0.750 | –0.050 | –0.600 | –0.058 |
| | (6.8) | | (6.5) | |
| Empleado | 0.186 | 0.012 | 0.402 | 0.039 |
| | (1.6) | | (3.6) | |
| Propietario de vivienda | 0.520 | 0.035 | 0.522 | 0.051 |
| | (4.7) | | (5.3) | |
| Log verosimilitud | –430.7 | | –526.0 | |
| Estadístico ji cuadrada (H0: Todos los coeficientes, excepto la constante, son iguales a cero) | 408 | | 602 | |
| Número de observaciones | 2 025 | | 2 091 | |
| Porcentaje de la muestra con predicciones correctas | 91 | | 90 | |

* Las cifras entre paréntesis son estadísticos *t*.

Fuente: John P. Caskey y Andrew Peterson, “Who Has a Bank Account and Who Doesn’t: 1977 and 1989”, *Research Working Paper* 93-10, Federal Reserve Bank of Kansas City, octubre de 1993.

15.18. *Estudio Monte Carlo.* Como ayuda para entender el modelo probit, William Becker y Donald Waldman supusieron lo siguiente:*

$$E(Y \mid X) = -1 + 3X$$

Así, con $Y_i = -1 + 3X + \varepsilon_i$, donde $\varepsilon_i$ se supone normal estándar (es decir, media cero y varianza unitaria), generaron una muestra de 35 observaciones, como se muestra en la tabla 15.26.

*a*) De la información sobre $Y$ y $X$ de esta tabla, ¿puede estimar un MLP? Recuerde que la verdadera $E(Y \mid X) = -1 + 3X$.

*b*) Con $X = 0.48$, estime $E(Y \mid X) = 0.48)$ y compárela con la verdadera $E(Y \mid X) = 0.48)$. Note que $\bar{X} = 0.48$.

*c*) Con la información sobre $Y^*$ y $X$ de la tabla 15.26, estime un modelo probit. Utilice el paquete estadístico que desee. El modelo probit estimado por los autores es el siguiente:

$$\hat{Y}_i^* = -0.969 + 2.764X_i$$

Encuentre la $P(Y^* = 1 \mid X = 0.48)$, es decir, $P(Y_1 > 0 \mid X = 0.48)$. Vea si la respuesta coincide con la respuesta de 0.64 de los autores.

*d*) La desviación estándar muestral de los valores $X$ de la tabla 15.26 es 0.31. ¿Cuál es el cambio proyectado en la probabilidad si $X$ es una desviación estándar por encima del valor de la media, es decir, cuál es $P(Y^* = 1 \mid X = 0.79)$? La respuesta de los autores es 0.25.

**TABLA 15.26**

**Conjunto de datos hipotéticos generados por el modelo $Y = -1 + 3X + \varepsilon$ y $Y^* = 1$, si $Y > 0$**

Fuente: William E. Becker y Donald M. Waldman, "A Graphical Interpretation of Probit Coefficients", *Journal of Economic Education*, otoño de 1989, tabla 1, p. 373.

| *Y* | *Y** | *X* | *Y* | *Y** | *X* |
|---|---|---|---|---|---|
| –0.3786 | 0 | 0.29 | –0.3753 | 0 | 0.56 |
| 1.1974 | 1 | 0.59 | 1.9701 | 1 | 0.61 |
| –0.4648 | 0 | 0.14 | –0.4054 | 0 | 0.17 |
| 1.1400 | 1 | 0.81 | 2.4416 | 1 | 0.89 |
| 0.3188 | 1 | 0.35 | 0.8150 | 1 | 0.65 |
| 2.2013 | 1 | 1.00 | –0.1223 | 0 | 0.23 |
| 2.4473 | 1 | 0.80 | 0.1428 | 1 | 0.26 |
| 0.1153 | 1 | 0.40 | –0.6681 | 0 | 0.64 |
| 0.4110 | 1 | 0.07 | 1.8286 | 1 | 0.67 |
| 2.6950 | 1 | 0.87 | –0.6459 | 0 | 0.26 |
| 2.2009 | 1 | 0.98 | 2.9784 | 1 | 0.63 |
| 0.6389 | 1 | 0.28 | –2.3326 | 0 | 0.09 |
| 4.3192 | 1 | 0.99 | 0.8056 | 1 | 0.54 |
| –1.9906 | 0 | 0.04 | –0.8983 | 0 | 0.74 |
| –0.9021 | 0 | 0.37 | –0.2355 | 0 | 0.17 |
| 0.9433 | 1 | 0.94 | 1.1429 | 1 | 0.57 |
| –3.2235 | 0 | 0.04 | –0.2965 | 0 | 0.18 |
| 0.1690 | 1 | 0.07 | | | |

* William E. Becker y Donald M. Waldman, "A Graphical Interpretation of Probit Coefficients", *Journal of Economic Education*, vol. 20, núm. 4, otoño de 1989, pp. 371-378.

15.19. La tabla 15.27 del sitio web del libro de texto presenta datos de 2 000 mujeres en relación con el trabajo (1 = la mujer trabaja, 0 = no trabaja), edad, estado civil (1 = casada, 0 = otros), número de hijos y nivel de escolaridad (número de años de escolaridad). De un total de 2 000 mujeres, se registraron 657 que no perciben salario.

*a*) Con estos datos, estime el modelo lineal de probabilidad (MLP).

*b*) Con los mismos datos, estime un modelo logit y obtenga los efectos marginales de las diferentes variables.

*c*) Repita *b*) con el modelo probit.

*d*) ¿Qué modelo elegiría? ¿Por qué?

15.20. Para el ejemplo de los fumadores analizado en el texto (sección 15.10), descargue los datos del sitio web del libro de texto que se presentan en la tabla 15.28. Vea si el producto de escolaridad e ingreso (es decir, el efecto de la interacción) tiene algún efecto sobre la probabilidad de llegar a ser fumador.

15.21. Descargue el conjunto de datos Benign, que aparece en la tabla 15.29, del sitio web del libro de texto. La variable cáncer es dicótoma, donde 1 = tuvo cáncer mamario y 0 = no tuvo cáncer mamario.* Use las variables edad (= edad del sujeto), HIGD (= grado más alto de educación terminado), CHK (= 0 si el sujeto no se realizó exámenes médicos con regularidad y = 1 si el sujeto se realizó exámenes médicos con regularidad), AGPI (= edad cuando se embarazó por primera vez), abortos espontáneos ( = número de abortos espontáneos) y peso (= peso del sujeto) y ejecute una regresión logística para concluir si estas variables son estadísticamente útiles para pronosticar si una mujer se enfermará de cáncer mamario o no.

# Apéndice 15A

## 15A.1 Estimación de máxima verosimilitud de los modelos probit y logit para datos individuales (no agrupados)†

Como en el texto, suponga que nos interesa estimar la probabilidad de que un individuo sea propietario de una casa, dado el ingreso $X$ de dicha persona. Suponemos que esta probabilidad se expresa mediante la función logística (15.5.2), que se reproduce a continuación.

$$P_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \tag{1}$$

En realidad no observamos $P_i$, sino sólo el resultado $Y = 1$, si un individuo es propietario de una casa, y $Y = 0$ si no lo es.

Como cada $Y_i$ es una variable aleatoria Bernoulli, se expresa

$$\Pr(Y_i = 1) = P_i \tag{2}$$

$$\Pr(Y_i = 0) = (1 - P_i) \tag{3}$$

* Se proporcionan datos sobre 50 mujeres a quienes se les diagnosticó enfermedad mamaria benigna y 150 controles que corresponden a la edad, con tres controles por caso. Los entrevistadores profesionales administraron un cuestionario estructurado estandarizado para recopilar información de cada sujeto del estudio (véase Pastides *et al.*, 1983, y Pastides *et al.*, 1985).

† El siguiente análisis se basa en gran medida en John Neter, Michael H. Kutner, Christopher J. Nachsteim y William Wasserman, *Applied Linear Statistical Models*, 4a. ed., Irwin, 1996, pp. 573–574.

Suponga que tenemos una *muestra aleatoria* de $n$ observaciones. Sea la función $f_i(Y_i)$ tal que denote la probabilidad de que $Y_i = 1$ o 0; la probabilidad conjunta de observar los $n$ valores $Y$, es decir, $f(Y_1, Y_2, \ldots, Y_n)$, se expresa como:

$$f(Y_1, Y_2, \ldots, Y_n) = \prod_1^n f_i(Y_i) = \prod_1^n P_i^{Y_i}(1 - P_i)^{1-Y_i} \tag{4}$$

donde $\prod$ es el operador producto; observe que escribimos la función de densidad de probabilidades conjuntas como producto de las funciones de densidad individuales, pues cada $Y_i$ se obtiene de manera independiente y cada $Y_i$ tiene la misma función de densidad (logística). La probabilidad conjunta dada en la ecuación (4) se conoce como **función de verosimilitud (FV)**.

Es un poco difícil manipular la ecuación (4). Pero si tomamos su logaritmo natural, obtenemos lo que se conoce como **función log de verosimilitud (FLV)**:

$$\begin{aligned}
\ln f(Y_1, Y_2, \ldots, Y_n) &= \sum_1^n [Y_i \ln P_i + (1 - Y_i) \ln (1 - P_i)] \\
&= \sum_1^n [Y_i \ln P_i - Y_i \ln (1 - P_i) + \ln (1 - P_i)] \qquad (5) \\
&= \sum_1^n \left[Y_i \ln\left(\frac{P_i}{1 - P_i}\right)\right] + \sum_1^n \ln (1 - P_i)
\end{aligned}$$

De (1) resulta fácil verificar que

$$(1 - P_i) = \frac{1}{1 + e^{\beta_1 + \beta_2 X_i}} \tag{6}$$

así como

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 X_i \tag{7}$$

Mediante (6) y (7) expresamos la FLV(5) como:

$$\ln f(Y_1, Y_2, \ldots, Y_n) = \sum_1^n Y_i(\beta_1 + \beta_2 X_i) - \sum_1^n \ln\left[1 + e^{(\beta_1 + \beta_2 X_i)}\right] \tag{8}$$

Como se observa de (8), la función log de verosimilitud es una función de los parámetros $\beta_1$ y $\beta_2$, pues se conocen las $X_i$.

En MV, el objetivo consiste en maximizar la FV (o la FLV), es decir, en obtener los valores de los parámetros desconocidos de forma que la probabilidad de observar las $Y$ dadas sea tan grande (máximo) como sea posible. Con este propósito, diferenciamos (8) parcialmente respecto de cada incógnita, igualamos las expresiones resultantes a cero y resolvemos las expresiones así obtenidas. Luego aplicamos la condición de maximización de segundo orden a fin de verificar que los valores de los parámetros obtenidos en verdad maximicen la FV.

Por tanto, hay que diferenciar (8) respecto de $\beta_1$ y $\beta_2$, y proceder como se indicó. Como puede observar de inmediato, las expresiones resultantes se convierten en no lineales en grado sumo respecto de los parámetros, por lo que no se pueden obtener soluciones explícitas. Ésta es la razón por la que tendremos que utilizar uno de los métodos de estimación no lineal analizados en el capítulo anterior con el objeto de obtener soluciones numéricas. Una vez que se obtienen los valores numéricos de $\beta_1$ y $\beta_2$, la ecuación (1) se estima con facilidad.

El procedimiento de MV para el modelo probit es semejante al del modelo logit, excepto que en (1) utilizamos la FDA normal en vez de la FDA logística. La expresión resultante se complica un poco, pero la idea general es la misma. Así que no profundizaremos más.

Capítulo 16

# Modelos de regresión con datos de panel

En el capítulo 1 analizamos con brevedad los tipos de datos que, por lo general, están disponibles para el análisis empírico, a saber: **de series de tiempo, de corte transversal** y **de panel**. En los datos de series de tiempo se observan los valores de una o más variables durante un periodo (por ejemplo, el PIB durante varios trimestres o varios años). En los datos de corte transversal, se recopilan valores de una o más variables para varias unidades muestrales, o entidades, en el mismo punto en el tiempo (por ejemplo, la tasa de criminalidad en 50 estados de Estados Unidos en un año determinado). En los datos de panel, la misma unidad de corte transversal (una familia, una empresa o un estado) se estudia a lo largo del tiempo. En resumen, en los datos de panel está la dimensión del espacio y la del tiempo.

Ya vimos un ejemplo de lo anterior en la tabla 1.1, la cual suministra datos sobre los huevos producidos y sus precios en 50 estados de Estados Unidos en 1990 y 1991. En cualquier año dado, los datos de los huevos y su precio representan una muestra transversal. Para un estado dado existen dos observaciones de series de tiempo: para los huevos y para su precio. En consecuencia, se tienen en total 100 observaciones (agrupadas) sobre los huevos producidos y sus precios.

En la tabla 1.2 presentamos otro ejemplo de los datos de panel. Esta tabla contiene datos sobre inversión, valor de la empresa y capital accionario de cuatro empresas de 1935 a 1954. Los datos de cada empresa correspondientes al periodo 1935-1954 constituyen *datos de series de tiempo,* con 20 observaciones; los datos de las cuatro empresas para un año determinado son *datos de corte transversal,* con sólo cuatro observaciones; y los datos de todas las empresas para todos los años son *datos de panel,* con un total de 80 observaciones.

Existen otros nombres para los datos de panel, como **datos agrupados** (agrupamiento de observaciones de series de tiempo y de corte transversal); **combinación de datos de series de tiempo y de corte transversal**; **datos de micropanel**; **datos longitudinales** (un estudio a lo largo del tiempo de una variable o grupo de sujetos); **análisis de historia de sucesos** (por ejemplo, el estudio del movimiento de sujetos a lo largo del tiempo y a través de sucesivos estados o condiciones); **análisis de generaciones** (por ejemplo, dar seguimiento a la trayectoria profesional de los egresados en 1965 de una escuela de administración). Si bien hay variaciones sutiles, todos estos nombres en esencia tienen una connotación de movimiento de unidades de corte transversal a lo largo del tiempo. Por consiguiente, de forma genérica aplicaremos el término datos de panel para incluir uno o más de tales términos. Asimismo, denominaremos modelos de regresión con datos de **panel a los modelos de regresión** que se basan en tales datos.

Cada vez son más comunes los datos de panel en la investigación económica. Algunos conjuntos de datos de panel bien conocidos son:

1. El **estudio de panel de la dinámica del ingreso (EPDI)** llevado a cabo por el Instituto de Investigación Social de la Universidad de Michigan. Iniciado en 1968, cada año el Instituto recopila datos sobre casi 5 000 familias respecto de diversas variables socioeconómicas y demográficas.

2. La Oficina del Censo, perteneciente al Departamento de Comercio de Estados Unidos, lleva a cabo un estudio similar al EPDI, llamado **Estudio del ingreso y participación en programas (EIPP)**. Cuatro veces al año se entrevista a los participantes para conocer su situación económica.
3. El **panel alemán socioeconómico (PALSOE)** estudió a 1 761 individuos cada año entre 1984 y 2002. Se recopiló información sobre año de nacimiento, sexo, satisfacción con la vida, estado civil, percepciones laborales individuales y horas de trabajo anuales de cada sujeto del estudio de 1984 a 2002.

Asimismo, existen muchos otros estudios que llevan a cabo varias oficinas gubernamentales, como:

Estudio de familias, ingresos y dinámica del trabajo en Australia (HILDA, por sus siglas en inglés).
Estudio de panel de familias británicas (BHPS, por sus siglas en inglés).
Estudio de panel de mano de obra e ingreso en Corea (KLIPS, por sus siglas en inglés).

Desde el principio conviene hacer una advertencia: el tema de la regresión con datos de panel es muy amplio, y parte de las matemáticas y las estadísticas que implica son muy complejas. Sólo pretendemos abarcar algunos aspectos esenciales de los modelos de regresión con datos de panel, y dejamos los detalles para consulta en las referencias.[1] Pero advertimos que algunas referencias son excesivamente técnicas. Por fortuna, existen paquetes de software accesibles, como LIMDEP, PC-GIVE, SAS, STATA, SHAZAM y *EViews*, entre otros; estos programas facilitan en gran medida la tarea de implementar realmente las regresiones con datos de panel.

## 16.1 ¿Por qué datos de panel?

¿Cuáles son las ventajas de los datos de panel respecto de los datos de corte transversal o de series de tiempo? Baltagi menciona las siguientes ventajas de los datos de panel:[2]

1. Como los datos de panel se refieren a individuos, empresas, estados, países, etc., a lo largo del tiempo, lo más seguro es la presencia de *heterogeneidad* en estas unidades. Las técnicas de estimación de datos de panel toman en cuenta de manera explícita tal heterogeneidad, al permitir la existencia de variables específicas por sujeto, como veremos en breve. Utilizamos el término **sujeto** en un sentido genérico para incluir microunidades como individuos, empresas, estados y países.
2. Al combinar las series de tiempo de las observaciones de corte transversal, los datos de panel proporcionan "una mayor cantidad de datos informativos, más variabilidad, menos colinealidad entre variables, más grados de libertad y una mayor eficiencia".
3. Al estudiar las observaciones en unidades de corte transversal repetidas, los datos de panel resultan más adecuados para estudiar la dinámica del cambio. Los conjuntos de datos respecto del desempleo, la rotación en el trabajo y la movilidad laboral se estudian mejor con datos de panel.
4. Los datos de panel detectan y miden mejor los efectos que sencillamente ni siquiera se observan en datos puramente de corte transversal o de series de tiempo. Por ejemplo, los efectos

[1] Algunas referencias son G. Chamberlain, "Panel Data", en *Handbook of Econometrics*, vol. II; Z. Griliches y M.D. Intriligator (eds.), North-Holland, 1984, cap. 22; C. Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986; G.C. Judge, R.C. Hill, W.E. Griffiths, H. Lutkepohl y T.C. Lee, *Introduction to the Theory and Practice of Econometrics*, 2a. ed., John Wiley & Sons, Nueva York, 1985, cap. 11; W.H. Greene, *Econometric Analysis*, 6a. ed., Prentice-Hall, Englewood Cliffs, Nueva Jersey, 2008, cap. 19; Badi H. Baltagi, *Econometric Analysis of Panel Data*, John Wiley and Sons, Nueva York, 1995, y J.M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts, 1999. Para un análisis detallado del tema con aplicaciones empíricas, véase Edward W. Frees, *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences,* Cambridge University Press, Nueva York, 2004.

[2] Baltagi, *op. cit.*, pp. 3-6.

de las leyes concernientes al salario mínimo sobre el empleo y los salarios se estudian mejor si incluimos oleadas sucesivas de incrementos a los salarios mínimos en los salarios mínimos estatales y/o federales.

5. Los datos de panel permiten estudiar modelos de comportamiento más complejos. Por ejemplo, fenómenos como las economías de escala y el cambio tecnológico son más maniobrables con los datos de panel que con los datos puramente de corte transversal o de series de tiempo.
6. Al hacer disponibles datos para varios miles de unidades, los datos de panel reducen el sesgo posible si se agregan individuos o empresas en conjuntos numerosos.

En resumen, los datos de panel enriquecen el análisis empírico de manera que no sería posible con sólo datos de corte transversal o de series de tiempo. Con lo anterior no sugerimos que se eliminan todos los problemas con los modelos de datos de panel. Analizaremos dichos problemas después de ver algo de teoría y analizar algunos ejemplos.

## 16.2 Datos de panel: un ejemplo ilustrativo

Con el objeto de preparar el terreno, presentemos un ejemplo concreto. Considere los datos de la tabla 16.1 del sitio web del libro de texto, los cuales se recopilaron originalmente por el profesor Moshe Kim y se tomaron del libro de William Greene.[3] Los datos analizan los costos de seis líneas de aviación comercial de 1970 a 1984, para un total de 90 observaciones de datos de panel.

Las variables se definen como sigue: $I$ = identificación de la aerolínea; $T$ = identificación del año; $Q$ = producción, como ingresos por milla por pasajero, un índice; $C$ = costo total, en 1 000 dólares; $PF$ = precio del combustible; y $LF$ = factor de carga, la utilización promedio de la capacidad de la flotilla.

Suponga que deseamos averiguar cómo se comporta el costo total (C) en relación con la producción ($Q$), el precio del combustible ($PF$) y el factor de carga ($LF$). En resumen, deseamos estimar la función de costos de la aerolínea.

¿Cómo estimamos esta función? Desde luego, podemos estimar la función de costo de cada aerolínea con los datos de 1970 a 1984 (es decir, una regresión de series de tiempo). Esto se logra con el procedimiento acostumbrado de mínimos cuadrados ordinarios (MCO). En total habrá seis funciones de costo, una para cada aerolínea. Sin embargo, olvidaríamos así la información sobre las otras aerolíneas que operan en el mismo entorno (de regulación).

También podemos estimar una función de costo de corte transversal (es decir, una regresión de corte transversal). En total habrá 15 regresiones de corte transversal, una por año. Pero esto no tendría mucho sentido en el presente contexto, pues sólo se tienen seis observaciones por año y hay tres variables explicativas (más el término del intercepto); habrá muy pocos grados de libertad para realizar un análisis significativo. Además, no se "explota" el carácter de panel de los datos.

Por cierto, los datos de panel del ejemplo se denominan **panel balanceado**; se dice que un panel es balanceado si cada sujeto (empresa, individuos, etc.) tiene el mismo número de observaciones. Si cada entidad tiene un número diferente de observaciones, se tiene un **panel desbalanceado**. En la mayor parte de este capítulo trabajaremos con paneles balanceados. En la bibliografía de datos de panel también se mencionan los términos **panel corto** y **panel largo**. En un panel corto, el número de sujetos de corte transversal, $N$, es mayor que el número de periodos, $T$. En un panel largo, $T$ es mayor que $N$. Como explicaremos más adelante, las técnicas de estimación dependen de que se cuente con un panel corto o uno largo.

Entonces, ¿qué opciones hay? Existen cuatro posibilidades.

1. **Modelo de MCO agrupados.** Tan sólo se agrupan las 90 observaciones y se estima una "gran" regresión, sin atender la naturaleza de corte transversal y de series de tiempo de los datos.

[3] William H. Greene, *Econometric Analysis,* 6a. ed., 2008. Los datos se localizan en http://pages.stern.nyu.edu/~wgreen/Text/econometricanalysis.htm.

2. **Modelo de mínimos cuadrados con variable dicótoma (MCVD) de efectos fijos.** Aquí se agrupan las 90 observaciones, pero se permite que cada unidad de corte transversal (es decir, cada aerolínea en este ejemplo) tenga su propia variable dicótoma (intercepto).
3. **Modelo de efectos fijos dentro del grupo.** En este caso también se agrupan las 90 observaciones, pero por cada aerolínea expresamos cada variable como una desviación de su valor medio y luego estimamos una regresión de MCO sobre los valores *corregidos por la media* o "sin media".
4. **Modelo de efectos aleatorios (MEFA).** A diferencia del modelo de MCVD, en el que se permite que cada aerolínea tenga su propio valor de intercepto (fijo), suponemos que los valores del intercepto son una extracción aleatoria de una población mucho mayor de aerolíneas.

A continuación analizamos cada método con los datos de la tabla 16.1. (Véase el sitio web del libro de texto.)

## 16.3 Modelo de regresión con MCO agrupados o de coeficientes constantes

Considere el siguiente modelo:

$$C_{it} = \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \qquad (16.3.1)$$
$$i = 1, 2, \ldots, 6$$
$$t = 1, 2, \ldots, 15$$

donde $i$ es el $i$-ésimo sujeto y $t$ es el periodo para las variables que se definieron antes. Elegimos la función lineal de costo para efectos ilustrativos, pero en el ejercicio 16.10 se pedirá estimar una función log-lineal, o doble-log, en cuyo caso los coeficientes de la pendiente darán las estimaciones de elasticidad.

Observe que agrupamos las 90 observaciones en su totalidad, pero tenga en cuenta que suponemos que los coeficientes de regresión son iguales para todas las aerolíneas. Es decir, no hay distinción entre ellas: una aerolínea es tan buena como otra, supuesto quizá difícil de sostener.

Se supone que las variables explicativas no son estocásticas. Si lo son, no están correlacionadas con el término de error. En ocasiones se supone que las variables explicativas son **estrictamente exógenas**. *Se dice que una variable es estrictamente exógena si no depende de los valores actuales, pasados y futuros del término de error* $u_{it}$.

También se supone que el término de error es $u_{it} \sim iid(0, \sigma_u^2)$, es decir, distribuido de manera independiente e idéntica, con media cero y varianza constante. Para efectos de la prueba de hipótesis, podemos suponer que el término de error está distribuido normalmente. Observe la notación con doble subíndice de la ecuación (16.3.1), la cual se explica por sí misma.

Presentemos primero los resultados de la ecuación estimada (16.3.1) y luego analizaremos los problemas que presenta este modelo. Los resultados de la regresión basados en *EViews,* versión 6, se presentan en la tabla 16.2.

Si examina los resultados de la **regresión agrupada** y aplica los criterios convencionales, verá que todos los coeficientes de regresión no sólo son muy significativos estadísticamente, sino que también concuerdan con las expectativas previas y que el valor de $R^2$ es muy alto. El único "pelo en la sopa" es que el estadístico estimado de Durbin-Watson es muy bajo, lo que indica una posible autocorrelación o correlación espacial de los datos. Por supuesto, como bien sabemos, un valor bajo Durbin-Watson también puede deberse a errores de especificación.

El problema principal de este modelo es que no distingue entre las diferentes aerolíneas ni indica si la respuesta de costo total a las variables explicativas a través del tiempo es la misma para todas las aerolíneas. En otras palabras, si agrupamos diferentes aerolíneas en diferentes periodos se *oculta* la **heterogeneidad** (individualidad o singularidad) que puede existir entre las aerolíneas. Otra forma de plantear esto es que la individualidad de cada sujeto se subsume en el término de perturbación $u_{it}$. En consecuencia, es muy posible que el término de error se correlacione con algunas regresoras del modelo. En ese caso, los coeficientes estimados en la ecuación (16.3.1) pueden estar sesgados, además de ser inconsistentes. Recuerde que un supuesto impor-

**TABLA 16.2**

Variable dependiente: *C*
Método: Mínimos cuadrados
Observaciones incluidas: 90

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* (intercepto) | 1158559. | 360592.7 | 3.212930 | 0.0018 |
| *Q* | 2026114. | 61806.95 | 32.78134 | 0.0000 |
| PF | 1.225348 | 0.103722 | 11.81380 | 0.0000 |
| LF | -3065753. | 696327.3 | -4.402747 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.946093 | Media de la variable dependiente | 1122524. |
| *R* cuadrada ajustada | 0.944213 | Desviación estándar de la variable dependiente | 1192075. |
| Error estándar de la regresión | 281559.5 | Estadístico *F* | 503.1176 |
| Suma de cuadrados residual | 6.82E+12 | Prob. (estadístico *F*) | 0.000000 |
| | | Durbin-Watson | 0.434162 |

tante del modelo clásico de regresión lineal es que no hay correlación entre las regresoras y el término de perturbación o error.

Para ver cómo el término de error se correlaciona con las regresoras, considere la siguiente revisión del modelo (16.3.1):

$$C_{it} = \beta_1 + \beta_2 PF_{it} + \beta_3 LF_{it} + \beta_4 M_{it} + u_{it} \qquad \textbf{(16.3.2)}$$

donde la variable adicional $M =$ filosofía de la administración o calidad de la administración. De las variables en la ecuación (16.3.2), sólo la variable $M$ es **invariante en el tiempo** (o **constante en el tiempo**) porque varía entre sujetos, pero es constante a través del tiempo para un sujeto (aerolínea) dado.

Aunque es invariante en el tiempo, la variable $M$ no puede observarse directamente y, por tanto, no se puede medir su contribución a la función de costo. Sin embargo, esto se consigue de manera indirecta si escribimos la ecuación (16.3.2) como

$$C_{it} = \beta_1 + \beta_2 PF_{it} + \beta_3 LF_{it} + \alpha_i + u_{it} \qquad \textbf{(16.3.3)}$$

donde $\alpha_i$, denominado **efecto no observado** o **de heterogeneidad**, refleja el efecto de $M$ sobre el costo. Observe que, para simplificar, sólo mostramos el efecto no observado de $M$ sobre el costo, pero en realidad puede haber más efectos no observados; por ejemplo, el carácter de la propiedad (pública o privada), si se trata de una empresa propiedad de una minoría, si el director general es hombre o mujer, etc. Aunque dichas variables pueden diferir entre sujetos (aerolíneas), es probable que no varíen para un sujeto dado durante el periodo de la muestra.

Como $\alpha_i$ no es directamente observable, ¿por qué no considerarlo aleatorio e incluirlo en el término de error $u_{it}$ y, por tanto, considerar el término de error compuesto $v_{it} = \alpha_i + u_{it}$? Ahora escribimos la ecuación (16.3.3) así:

$$C_{it} = \beta_1 + \beta_2 PF_{it} + \beta_3 LF_{it} + v_{it} \qquad \textbf{(16.3.4)}$$

Pero si el término $\alpha_i$ incluido en el término de error $v_{it}$ está correlacionado con cualquiera de las regresoras de la ecuación (16.3.4), hay una violación de un supuesto fundamental del modelo clásico de regresión lineal: que el término de error no está correlacionado con las regresoras. Como sabemos, en esta situación los estimados de MCO no sólo están sesgados, sino que también son inconsistentes.

Existe una posibilidad real de que el término $\alpha_i$ no observable esté correlacionado con una o más de las regresoras. Por ejemplo, la administración de una aerolínea puede actuar de manera inteligente y comprar contratos de futuro sobre el precio del combustible para evitar fluctuaciones bruscas de los precios. Esto tendría el efecto de reducir el costo de los servicios de la aerolínea. Como resultado de esta correlación, podemos demostrar que $\text{cov}(v_{it}, v_{is}) = \sigma_u^2$; $t \neq s$, que es diferente de cero, y, por tanto, la heterogeneidad (no observada) induce *autocorrelación* y habrá que prestarle atención. Más adelante indicaremos cómo manejar este problema.

En consecuencia, la pregunta es cómo se toman en cuenta los efectos no observables, o heterogeneidad, para obtener estimaciones consistentes y eficientes de los parámetros de las variables de interés primordial, que son producción, precio del combustible y factor de carga en este caso. El interés primordial quizá no se centre en obtener el efecto de las variables no observables porque éstas no cambian para un sujeto dado. Por esta razón, los efectos no observables, o heterogeneidad, se llaman **parámetros incómodos**. ¿Cómo proceder entonces? Lo veremos a continuación.

## 16.4 Modelo de mínimos cuadrados con variable dicótoma (MCVD) de efectos fijos

El modelo de mínimos cuadrados con variable dicótoma (MCVD) toma en cuenta la heterogeneidad entre sujetos porque permite que cada entidad tenga su propio valor del intercepto, como se muestra en el modelo (16.4.1). Una vez más, trabajaremos con el ejemplo de las aerolíneas.

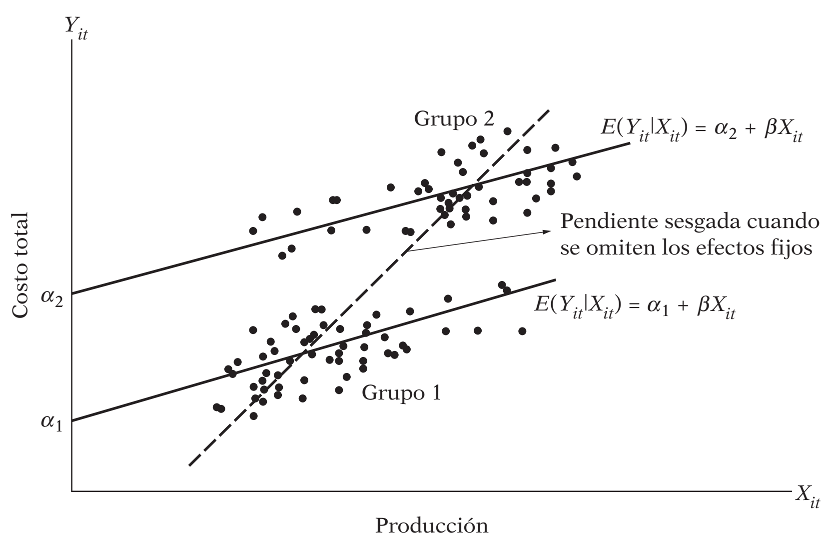
$$
\begin{aligned}
C_{it} &= \beta_{1i} + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \qquad \textbf{(16.4.1)}\\
i &= 1, 2 \ldots, 6\\
t &= 1, 2, \ldots, 15
\end{aligned}
$$

Observe que utilizamos el subíndice *i* en el término del intercepto para indicar que los interceptos de las seis aerolíneas pueden ser diferentes. Las diferencias quizá se deban a características especiales de cada aerolínea, como el estilo de administración, la filosofía de la empresa o el tipo de mercado que atiende cada aerolínea.

En la bibliografía, el modelo (16.4.1) se conoce como **modelo (regresión) de efectos fijos (MEF)**. El término "efectos fijos" se debe a que, aunque el intercepto puede diferir entre los sujetos (en este caso las seis aerolíneas), el intercepto de cada entidad no varía con el tiempo, es decir, es **invariante en el tiempo**. Observe que si el intercepto se escribiera $\beta_{1it}$, indicaría que el intercepto de cada entidad o individuo es **variable en el tiempo**. Cabe señalar que el MEF dado en la ecuación (16.4.1) supone que los coeficientes (de las pendientes) de las regresoras no varían según los individuos ni a través del tiempo.

Antes de seguir adelante, es útil visualizar la diferencia entre el modelo de regresión agrupada y el modelo de MCVD. Para simplificar, suponga que deseamos efectuar una regresión del costo total sólo sobre la producción. En la figura 16.1 se muestra esta función de costo estimada para dos aerolíneas por separado, así como la función de costo si agrupamos los datos de las dos em-

**FIGURA 16.1**
Sesgo por omitir los efectos fijos.



presas; esto equivale a no tomar en cuenta los efectos fijos.[4] En la figura 16.1 se observa cómo la regresión agrupada sesga la estimación de la pendiente.

¿Cómo se permite en realidad que el intercepto (de efecto fijo) varíe entre las aerolíneas? Se realiza con facilidad mediante la técnica de las variables dicótomas que explicamos en el capítulo 9, en particular las **variables dicótomas con intercepto diferencial**. Ahora expresamos así la ecuación (16.4.1):

$$C_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \alpha_5 D_{5i} + \alpha_6 D_{6i} + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \tag{16.4.2}$$

donde $D_{2i} = 1$ si la observación corresponde a la aerolínea 2, y 0 en otro caso; $D_{3i} = 1$ si la observación es de la aerolínea 3, y 0 en otro caso; y así sucesivamente. Como se trata de seis aerolíneas, sólo utilizamos cinco variables dicótomas para evitar caer en la **trampa de la variable dicótoma** (es decir, una situación de colinealidad perfecta). En este caso, la aerolínea 1 se considera la categoría base o de referencia. Desde luego, podemos elegir cualquier aerolínea como punto de referencia. Como resultado, el intercepto $\alpha_1$ es el valor del intercepto de la aerolínea 1, y los demás coeficientes $\alpha$ representan el grado en que los valores de los interceptos de las demás aerolíneas difieren del valor del intercepto de la primera aerolínea. Así, $\alpha_2$ indica por cuánto difiere de $\alpha_1$ el valor del intercepto de la segunda aerolínea. La suma $(\alpha_1 + \alpha_2)$ da el valor real del intercepto de la aerolínea 2. Los valores de los interceptos de las demás aerolíneas se calculan del mismo modo. *Recuerde que si desea introducir una variable dicótoma para cada aerolínea, es necesario omitir el intercepto (común); de lo contrario, caerá en la trampa de la variable dicótoma.*

Los resultados del modelo (16.4.2) para estos datos se presentan en la tabla 16.3.

Lo primero que debe notarse en estos resultados es que todos los coeficientes de los interceptos diferenciales son muy significativos estadísticamente en lo individual, lo cual indica que tal vez las seis aerolíneas son heterogéneas y, por tanto, los resultados de la regresión agrupada presentados en la tabla 16.2 son dudosos. Los valores de los coeficientes de las pendientes de las tablas 16.2 y 16.3 también son diferentes, lo que una vez más arroja dudas sobre los resultados de la tabla 16.2. Al parecer, el modelo (16.4.1) es mejor que el modelo (16.3.1). A propósito, observe que los MCO aplicados a un modelo de efectos fijos producen estimadores que se llaman **estimadores de efectos fijos**.

**TABLA 16.3**

Variable dependiente: CT
Método: Mínimos cuadrados
Muestra: 1-90
Observaciones incluidas: 90

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| $C$ $(=\alpha_1)$ | -131236.0 | 350777.1 | -0.374129 | 0.7093 |
| *Q* | 3319023. | 171354.1 | 19.36939 | 0.0000 |
| *PF* | 0.773071 | 0.097319 | 7.943676 | 0.0000 |
| *LF* | -3797368. | 613773.1 | -6.186924 | 0.0000 |
| DIC2 | 601733.2 | 100895.7 | 5.963913 | 0.0000 |
| DIC3 | 1337180. | 186171.0 | 7.182538 | 0.0000 |
| DIC4 | 1777592. | 213162.9 | 8.339126 | 0.0000 |
| DIC5 | 1828252. | 231229.7 | 7.906651 | 0.0000 |
| DIC6 | 1706474. | 228300.9 | 7.474672 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.971642 | Media de la variable dependiente | 1122524. |
| *R* cuadrada ajustada | 0.968841 | Desviación estándar de la variable dependiente | 1192075. |
| Error estándar de la regresión | 210422.8 | Estadístico *F* | 346.9188 |
| Suma de cuadrados residual | 3.59E+12 | Prob. (estadístico *F*) | 0.000000 |
| Log verosimilitud | -1226.082 | Estadístico Durbin-Watson | 0.693288 |

[4] Adaptado de las notas inéditas de Alan Duncan.

Podemos proporcionar una prueba formal de los dos modelos. En relación con el modelo (16.4.1), el modelo (16.3.1) es un modelo *restringido* pues impone un intercepto común para todas las aerolíneas. En consecuencia, podemos utilizar la **prueba *F* restringida** analizada en el capítulo 8. Mediante la fórmula (8.6.10), el lector puede verificar que en el ejemplo presente el valor *F* es:

$$F = \frac{(0.971642 - 0.946093)/5}{(1 - 0.971642)/81} \approx 14.99$$

*Nota:* Los valores de $R^2$ restringida y no restringida se obtienen de las tablas 16.1 y 16.2. Observe también que el número de restricciones es 5 (¿por qué?)

La hipótesis nula en este caso es que todos los interceptos diferenciales son iguales a cero. El valor *F* calculado de 5 gl para el numerador y 81 gl para el denominador es muy significativo estadísticamente. Por tanto, se rechaza la hipótesis nula de que todos los interceptos (diferenciales) son iguales a cero. Si el valor *F* no fuera significativo estadísticamente, concluiríamos que no hay diferencias entre los interceptos de las seis aerolíneas. En este caso, habríamos agrupado las 90 observaciones, como en la regresión agrupada de la tabla 16.2.

El modelo (16.4.1) se conoce como modelo de **efectos fijos unidireccionales** porque permite que los interceptos varíen entre las aerolíneas. Pero también podemos permitir el **efecto tiempo** si creemos que la función de costo cambia con el tiempo a causa de otros factores, como los avances tecnológicos, variaciones en las políticas de regulación gubernamental o fiscal, así como otros efectos. Tales efectos temporales se toman en cuenta con facilidad si introducimos variables dicótomas temporales, una para cada año de 1970 a 1984. Como hay datos para 15 años, introducimos 14 variables dicótomas temporales (¿por qué?) y ampliamos el modelo (16.4.1) mediante la adición de estas variables. Si hacemos eso, el modelo resultante se llama **modelo de efectos fijos bidireccionales** porque tiene en cuenta los efectos tanto individuales como temporales.

En el presente ejemplo, si agregamos las variables dicótomas de tiempo, tendremos que estimar 23 coeficientes en total: el intercepto común, cinco variables dicótomas de las aerolíneas, 14 variables dicótomas de tiempo y tres coeficientes de pendientes. Como se ve, consumiremos varios grados de libertad. Además, si permitimos que los coeficientes de las pendientes difieran entre empresas, las cinco variables de empresas (aerolíneas) interactúan con cada una de las tres variables explicativas e introducen **coeficientes de pendientes diferenciales dicótomas**. Entonces habrá que estimar 15 coeficientes adicionales (interacciones de cinco variables dicótomas con tres variables explicativas). Como si esto fuera poco, si las 14 variables dicótomas de tiempo interactúan con las tres variables explicativas, tendremos un total de 42 coeficientes adicionales que deben estimarse. Como se aprecia, no queda ningún grado de libertad.

## Advertencia sobre el modelo de MCVD de efectos fijos

Como indica el análisis precedente, el modelo MCVD presenta algunos problemas que es necesario tener en cuenta:

*Primero*, si se introducen demasiadas variables dicótomas, puede presentarse el problema de los grados de libertad. Es decir, no habrá observaciones suficientes para un análisis estadístico significativo. *Segundo*, con tantas variables dicótomas en el modelo, tanto individuales como interactivas o multiplicativas, siempre está presente la posibilidad de la multicolinealidad, que puede dificultar la estimación precisa de uno o más parámetros.

*Tercero*, en algunas situaciones, es posible que el modelo de MCVD no identifique el efecto de las variables que no cambian con el tiempo. Suponga que deseamos estimar la función de salario de un grupo de trabajadores con datos de panel. Además del salario, una función de salario incluye edad, experiencia y nivel de escolaridad como variables explicativas. Suponga también que decidimos incluir género, color y origen étnico como variables adicionales del modelo. Como estas variables no cambian con el tiempo para cada individuo, el modelo MCVD quizá no identifique el impacto sobre los salarios de estas variables que no cambian con el tiempo. En otras palabras, los interceptos específicos de cada sujeto absorben toda la heterogeneidad que exista en las variables dependiente y explicativas. Por cierto, las variables que no cambian con el tiempo a veces se llaman **variables incómodas** o **variables ocultas**.

*Cuarto*, tenemos que reflexionar con mucho cuidado sobre el término de error $u_{it}$. Los resultados presentados en las ecuaciones (16.3.1) y (16.4.1) se basan en el supuesto de que el término de error se apega a los supuestos clásicos, a saber: $u_{it} \sim N(0, \sigma^2)$. Como el índice $i$ se refiere a las observaciones de corte transversal y $t$ a las observaciones de series de tiempo, quizá deba modificarse el supuesto clásico respecto de $u_{it}$. Existen diversas posibilidades:

1. Podemos suponer que la varianza del error es la misma para todas las unidades de corte transversal, o que la varianza del error es heteroscedástica.[5]
2. Para cada individuo, podemos suponer que no existe autocorrelación a través del tiempo. Entonces, en el ejemplo ilustrativo, podemos suponer que el término de error de la función de costo de la aerolínea 1 no está autocorrelacionado, o bien, que sí lo está, digamos con el tipo AR(1).
3. En un determinado tiempo, es posible que el término de error de la aerolínea 1 esté correlacionado con el término de error de la aerolínea 2, por ejemplo.[6] O bien, podemos suponer que no existe tal correlación.

Existen también otras permutaciones y combinaciones del término de error. Como es fácil darse cuenta, si se permite una o más de estas posibilidades, el análisis se complica mucho más. (El espacio disponible y las exigencias matemáticas impiden considerar todas las posibilidades. En las referencias de la nota 1 se analizan algunos de estos temas.) No obstante, algunos de estos problemas pueden resolverse con las opciones que analizamos en las siguientes dos secciones.

## 16.5 Estimador de efectos fijos dentro del grupo (DG)

Una forma de estimar una regresión agrupada es eliminar el efecto fijo, $\beta_{1i}$, expresando los valores de las variables dependiente y explicativas de cada aerolínea como desviaciones de sus respectivos valores medios. Así, para la aerolínea 1 obtenemos los valores muestrales medios de *TC, Q, PF* y *LF* ($\overline{TC}$, $\overline{Q}$, $\overline{PF}$ y $\overline{LF}$, respectivamente) y los restamos de los valores individuales de estas variables. Los valores resultantes se llaman valores *corregidos por la media,* o “sin media”. Hacemos esto con cada aerolínea y luego agrupamos todos los (90) valores corregidos por la media para ejecutar una regresión de MCO.

Si $tc_{it}$, $q_{it}$, $pf_{it}$ y $lf_{it}$ representan los valores corregidos por la media, efectuamos la siguiente regresión:

$$tc_{it} = \beta_2 q_{it} + \beta_3 pf_{it} + \beta_4 lf_{it} + u_{it} \tag{16.5.1}$$

donde $i = 1, 2, \ldots, 6$, y $t = 1, 2, \ldots, 15$. Observe que la ecuación (16.5.1) *no* tiene término de intercepto (¿por qué?).

De vuelta al ejemplo, obtenemos los resultados de la tabla 16.4. *Nota:* El prefijo DM significa que los valores están corregidos por la media o expresados como desviaciones de sus medias muestrales.

Observe la diferencia entre la regresión agrupada que presentamos en la tabla 16.2 y la de la tabla 16.4. La primera tan sólo omite la heterogeneidad entre las seis aerolíneas, mientras que la segunda la toma en cuenta no por el método de la variable dicótoma, sino que la elimina mediante la diferenciación de las observaciones de la muestra en torno a sus medias muestrales. La diferencia entre las dos es evidente, como se aprecia en la figura 16.2.

Se muestra que el estimador DG produce *estimaciones consistentes* de los coeficientes de pendiente, mientras que la regresión agrupada ordinaria tal vez no. Sin embargo, debe añadirse que

[5] STATA ofrece errores estándar corregidos por heteroscedasticidad en los modelos de regresión con datos de panel.

[6] Esto conduce al llamado **modelo de regresión aparentemente no relacionado** (**SURE**, por sus siglas en inglés), propuesto originalmente por Arnold Zellner. Véase A. Zellner, “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”, *Journal of the American Statistical Association,* vol. 57, 1962, pp. 348-368.

## TABLA 16.4

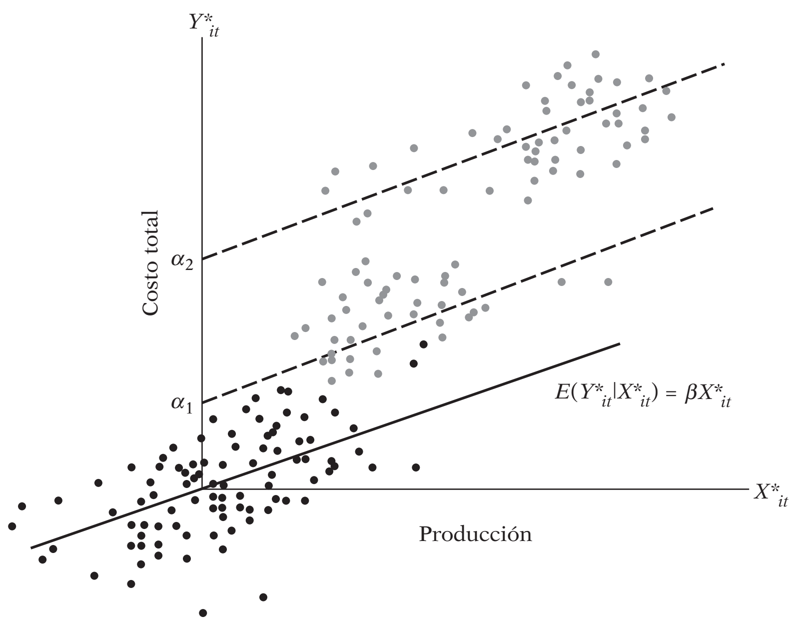
Variable dependiente: DMCT
Método: Mínimos cuadrados
Muestra: 1-90
Observaciones incluidas: 90

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| DMQ | 3319023. | 165339.8 | 20.07396 | 0.0000 |
| DMPF | 0.773071 | 0.093903 | 8.232630 | 0.0000 |
| DMLF | -3797368. | 592230.5 | -6.411976 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.929366 | Media de la variable dependiente | 2.59E-11 |
| *R* cuadrada ajustada | 0.927743 | Desviación estándar de la | 755325.8 |
| Error estándar de la regresión | 203037.2 | variable dependiente | 0.693287 |
| Suma de cuadrados residual | 3.59E+12 | Estadístico Durbin-Watson | |

**FIGURA 16.2**
Estimador dentro de grupos.

Fuente: Alan Duncan, "Cross-Section and Panel Data Econometrics", notas sin publicar para conferencia (adaptado).




los estimadores DG, aunque consistentes, son ineficientes (es decir, tienen varianzas grandes) en comparación con los resultados de la regresión agrupada ordinaria.[7] Observe que los coeficientes de pendiente de *Q, PF* y *LF* son idénticos en las tablas 16.3 y 16.4. *Esto es porque matemáticamente los dos modelos son idénticos.* A propósito, los coeficientes de regresión estimados por el método DG se llaman *estimadores DG.*

Una desventaja del estimador DG se explica con el siguiente modelo de regresión salarial:

$$W_{it} = \beta_{1i} + \beta_2\text{Experiencia}_{it} + \beta_3\text{Edad}_{it} + \beta_4\text{Género}_{it} + \beta_5\text{Educación}_{it} + \beta_6\text{Raza}_{it} \qquad \textbf{(16.5.2)}$$

En esta función de salario, las variables como género, educación y raza son invariantes en el tiempo. Si usamos estimadores DG, estas variables que no cambian con el tiempo se eliminarían

[7] La razón de esto es que, cuando las variables se expresan como desviaciones de sus valores medios, la variación en estos valores corregidos por la media es mucho más pequeña que la variación de los valores originales de las variables. En ese caso, la variación del término de perturbación $u_{it}$ puede ser relativamente grande, lo que produce errores estándar más altos de los coeficientes estimados.

(por la diferenciación). Como resultado, no se sabemos cómo reacciona el salario ante estas variables invariantes en el tiempo.[8] Pero es el precio que hay que pagar para evitar la correlación entre el término de error ($\alpha_i$ incluido en $v_{it}$) y las variables explicativas.

Otra desventaja del estimador DG es que "puede distorsionar los valores de los parámetros y desde luego eliminar los efectos de largo plazo".[9] *En general, cuando se diferencia una variable, se elimina el componente de largo plazo de esa variable.* Lo que queda es el valor de corto plazo de esa variable. Analizaremos este tema en detalle cuando expliquemos la econometría de series de tiempo, más adelante.

Al usar MCVD obtuvimos estimaciones directas de los interceptos de cada aerolínea. ¿Cómo podemos obtener las estimaciones de los interceptos con el método DG? En el ejemplo de las aerolíneas se obtienen como sigue:

$$\hat{\alpha}_i = \overline{C}_i - \hat{\beta}_2\overline{Q_i} - \hat{\beta}_3\overline{PF_i} - \hat{\beta}_4\overline{LF} \tag{16.5.3}$$

donde las barras sobre las variables denotan valores muestrales medios de las variables para la *i*-ésima aerolínea.

Es decir, obtuvimos el valor del intercepto de la *i*-ésima aerolínea al restar del valor medio de la variable dependiente los valores medios de las variables explicativas de esa aerolínea multiplicados por los coeficientes estimados de pendiente que se obtuvieron con los estimadores DG. Observe que los coeficientes estimados de pendiente permanecen igual para todas las aerolíneas, como se muestra en la tabla 16.4. Cabe señalar que el intercepto estimado en la ecuación (16.5.3) se asemeja al intercepto que estimamos en el modelo de regresión lineal estándar, que se presenta en la ecuación (7.4.21). Se deja al lector la tarea de calcular los interceptos de las seis aerolíneas de la manera mostrada y verificar que son iguales a los valores de los interceptos derivados en la tabla 16.3, salvo por errores de redondeo.

Observe que el intercepto estimado de cada aerolínea representa las características *específicas* de cada aerolínea, pero no podremos identificar estas características individualmente. Por consiguiente, el intercepto $\alpha_1$ de la aerolínea 1 representa la filosofía de la administración de esa aerolínea, la composición del consejo de administración, la personalidad del director general, el género del director general, etc. Todas estas características de heterogeneidad se integran al valor del intercepto. Como veremos más adelante, tales características se incluyen en el *modelo de efectos aleatorios.*

A propósito, debe señalarse que una alternativa al estimador DG es el **método de primeras diferencias**. En el método DG expresamos cada variable como una desviación del valor medio de dicha variable. En el método de primeras diferencias, por cada sujeto se obtienen diferencias sucesivas de las variables. Así, para la aerolínea 1 se resta la primera observación de *TC* de la segunda observación de *TC,* la segunda observación de *TC* de la tercera, y así sucesivamente. Hacemos lo mismo con cada variable restante y repetimos el proceso con las demás cinco aerolíneas. Después de este proceso tenemos sólo 14 observaciones por aerolínea, pues la primera observación no tiene valor previo. Como resultado, ahora tenemos 84 observaciones en lugar de las 90 originales. A continuación efectuamos la regresión de los valores de primeras diferencias de la variable *TC* sobre los valores de primeras diferencias de las variables explicativas como sigue:

$$\begin{aligned} \Delta TC_{it} &= \beta_2 \Delta Q_{it} + \beta_3 \Delta PF_{it} + \beta_4 \Delta LF_{it} + (u_{it} - u_{i,t-1}) \\ i &= 1, 2, \ldots, 6 \\ t &= 1, 2, \ldots, 84 \end{aligned} \tag{16.5.4}$$

donde $\Delta = (TC_{it} = TC_{i,\,t-1})$. Como señalamos en el capítulo 11, $\Delta$ se conoce como operador de primeras diferencias.[10]

[8] Esto también aplica al modelo MCVD.

[9] Dimitrios Asteriou y Stephen G. Hall, *Applied Econometrics: A Modern Approach,* Palgrave Macmillan, Nueva York, 2007, p. 347.

[10] Observe que la ecuación (16.5.3) no tiene término de intercepto (¿por qué?), pero podemos incluirlo si hay una variable de tendencia en el modelo original.

A propósito, observe que el término de perturbación original se sustituye con la diferencia entre los valores actuales y anteriores del término de perturbación. Si el término de perturbación original no está autocorrelacionado, la perturbación transformada *sí* lo está y, por tanto, presenta los tipos de problemas de estimación que analizamos en el capítulo 11. Sin embargo, si las variables explicativas son **estrictamente exógenas**, el estimador de primeras diferencias es insesgado, dados los valores de las variables explicativas. Asimismo, observe que el método de primeras diferencias tiene las mismas desventajas del método DG en el sentido de que las variables explicativas que no varían con el tiempo para un individuo se eliminan en la transformación de primeras diferencias.

Vale la pena señalar que los estimadores de primeras diferencias y de efectos fijos son iguales cuando sólo se tienen dos periodos, pero si hay más de dos periodos, estos estimadores difieren. Las razones de esto son muy complejas y el lector interesado puede consultar las referencias.[11] Se deja como ejercicio aplicar el método de primeras diferencias al ejemplo de las aerolíneas y comparar los resultados con los demás estimadores de efectos fijos.

## 16.6 Modelo de efectos aleatorios (MEFA)

Al comentar sobre los modelos de efectos fijos, o MCVD, Kmenta escribe:[12]

> Una pregunta obvia en conexión con el modelo de covarianza [es decir, MCVD] es si de verdad es necesario incluir variables dicótomas —con la consecuente pérdida de grados de libertad—. El fundamento del modelo de covarianza es que, al especificar el modelo de regresión, no hemos podido incluir variables explicativas relevantes que no varíen con el tiempo (y posiblemente otras que sí cambian con el tiempo, pero que tienen el mismo valor para todas las unidades de corte transversal), y que la inclusión de variables dicótomas es un *encubrimiento de nuestra ignorancia*.

Si las variables dicótomas en efecto reflejan una falta de conocimiento respecto del modelo (real), ¿por qué no expresar dicha ignorancia mediante el término de perturbación? Éste es precisamente el método propuesto por los partidarios del **modelo de componentes del error (MCE)**, o **modelo de efectos aleatorios (MEFA)**, que ilustraremos a continuación con la función de costo de las líneas aéreas.

La idea básica es comenzar con la ecuación (16.4.1):

$$TC_{it} = \beta_{1i} + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + u_{it} \tag{16.6.1}$$

En vez de considerar fija a $\beta_{1i}$, suponemos que es una variable aleatoria con un valor medio igual a $\beta_1$ (en este caso, sin subíndice $i$). Además, el valor del intercepto para una empresa individual se expresa como:

$$\beta_{1i} = \beta_1 + \varepsilon_i \tag{16.6.2}$$

donde $\varepsilon_i$ es un término de error aleatorio con valor medio igual a cero y varianza de $\sigma_\varepsilon^2$.

Lo que afirmamos en esencia es que las seis empresas de la muestra se tomaron de un universo mucho más grande de este tipo de compañías, que tienen una media común para el intercepto ($= \beta_1$) y que las diferencias individuales en los valores del intercepto de cada empresa se reflejan en el término de error $\varepsilon_i$.

Al sustituir (16.6.2) en (16.6.1) obtenemos:

$$\begin{aligned} TC_{it} &= \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + \varepsilon_i + u_{it} \\ &= \beta_1 + \beta_2 Q_{it} + \beta_3 PF_{it} + \beta_4 LF_{it} + w_{it} \end{aligned} \tag{16.6.3}$$

donde

$$w_{it} = \varepsilon_i + u_{it} \tag{16.6.4}$$

[11] Véase en particular Jeffrey M. Woolridge, *Econometric Analysis of Cross Section and Panel Data,* MIT Press, Cambridge, Massachusetts, 2002, pp. 279-283.

[12] Jan Kmenta, *Elements of Econometrics,* 2a. ed., Macmillan, Nueva York, 1986, p. 633.

El término de error compuesto $w_{it}$ consta de dos componentes, $\varepsilon_i$, componente de error de corte transversal o error específico del individuo, y $u_{it}$, la combinación del componente de error de series de tiempo y corte transversal, y que a veces se denomina **término idiosincrásico** porque varía en el corte transversal (es decir, el sujeto) así como en el tiempo. El *modelo de componentes del error* (MCE) debe su nombre a que el término de error compuesto consiste en dos (o más) componentes del error.

Los supuestos comunes en los que se basa el MCE son:

$$\begin{aligned} &\varepsilon_i \sim N(0, \sigma_\varepsilon^2) \\ &u_{it} \sim N\big(0, \sigma_u^2\big) \\ &E(\varepsilon_i u_{it}) = 0; \quad E(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j) \\ &E(u_{it} u_{is}) = E(u_{ij} u_{ij}) = E(u_{it} u_{js}) = 0 \quad (i \neq j; t \neq s) \end{aligned} \tag{16.6.5}$$

es decir, los componentes del error individuales no están correlacionados entre sí y no están autocorrelacionados en las unidades de series de tiempo ni en las de corte transversal. *También es muy importante observar que $w_{it}$ no está correlacionado con ninguna variable explicativa del modelo.* Como $\varepsilon_i$ es un componente de $w_{it}$, es posible que el segundo esté correlacionado con las variables explicativas. Si en efecto es así, el MCE producirá una estimación inconsistente de los coeficientes de regresión. En breve analizaremos la **prueba de Hausman**, que indica en una aplicación dada si $w_{it}$ está correlacionado con las variables explicativas, es decir, si el MCE es el modelo apropiado.

Observe con cuidado la diferencia entre el MEF y el MCE. En el MEF, cada unidad de corte transversal tiene su propio valor (fijo) de intercepto, en todos los valores de $N$ para $N$ unidades de corte transversal. Por otra parte, en el MCE, el intercepto (común) representa el *valor medio* de todos los interceptos (de corte transversal), y el componente de error $\varepsilon_i$ significa la desviación (aleatoria) del intercepto individual respecto de este valor medio. No obstante, tenga en cuenta que $\varepsilon_i$ no es directamente observable; esto se conoce como **variable inobservable** o **latente**.

Como resultado de los supuestos establecidos en (16.6.5), se deriva que:

$$E(w_{it}) = 0 \tag{16.6.6}$$

$$\text{var}(w_{it}) = \sigma_\varepsilon^2 + \sigma_u^2 \tag{16.6.7}$$

Ahora, si $\sigma_\varepsilon^2 = 0$, no hay diferencia entre los modelos (16.3.1) y (16.6.3), en cuyo caso tan sólo se agrupan todas las observaciones (de corte transversal y de series de tiempo) y se lleva a cabo la regresión agrupada, como hicimos en (16.3.1). Esto es válido porque en esta situación no hay efectos específicos del sujeto o porque todos se tomaron en cuenta en las variables explicativas.

Como muestra la ecuación (16.6.7), el término de error es homoscedástico. Sin embargo, puede demostrarse que $w_{it}$ y $w_{is}$ $(t \neq s)$ están correlacionados; es decir, los términos de error de una unidad de corte transversal dada en dos puntos en el tiempo están correlacionados. El coeficiente de correlación, corr($w_{it}$, $w_{is}$), es el siguiente:

$$\rho = \text{corr}(w_{it}, w_{is}) = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_u^2}; \quad t \neq s \tag{16.6.8}$$

Observe dos características especiales del coeficiente de correlación anterior. *Primera*, para cualquier unidad de corte transversal dada, el valor de la correlación entre los términos de error en dos momentos sigue siendo el mismo, sin importar la distancia entre los dos periodos, como resulta evidente de (16.6.8). Esto representa un fuerte contraste con el esquema de primer orden [AR(1)] que analizamos en el capítulo 12, donde descubrimos que la correlación entre periodos decrece con el tiempo. *Segunda*, la estructura de correlación dada en (16.6.8) sigue siendo la misma para todas las unidades de corte transversal; es decir, es idéntica para todos los sujetos.

Si no tomamos en cuenta esta estructura de correlación y calculamos (16.6.3) mediante MCO, los estimadores resultantes serán ineficientes. El método más adecuado en este caso es el de **mínimos cuadrados generalizados (MCG)**.

**TABLA 16.5**

Variable dependiente: CT
Método: Panel EMCG (efectos aleatorios de corte transversal)
Muestra: 1-15
Periodos incluidos: 15
Secciones de corte transversal incluidas: 6
Total de observaciones de panel (balanceado): 90
Estimador Swamy y Arora de varianzas de componentes

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | 107429.3 | 303966.2 | 3.534251 | 0.0007 |
| *Q* | 2288588. | 88172.77 | 25.95572 | 0.0000 |
| *PF* | 1.123591 | 0.083298 | 13.48877 | 0.0000 |
| *LF* | -3084994. | 584373.2 | -5.279151 | 0.0000 |

Especificación de efectos

| | D.E. | Rho |
|---|---|---|
| Aleatorios de corte transversal | 107411.2 | 0.2067 |
| Aleatorios idiosincrásicos | 210422.8 | 0.7933 |

| | Empresa | Efecto |
|---|---|---|
| 1 | 1.000000 | -270615.0 |
| 2 | 2.000000 | -87061.32 |
| 3 | 3.000000 | -21338.40 |
| 4 | 4.000000 | 187142.9 |
| 5 | 5.000000 | 134488.9 |
| 6 | 6.000000 | 57383.00 |

Dentro del presente contexto, no analizaremos las matemáticas de MCG, debido a su complejidad.[13] En vista de que el software estadístico más moderno cuenta ya con rutinas para estimar los MCE (así como los MEF), sólo presentaremos los resultados del ejemplo ilustrativo. Pero antes de hacerlo, conviene notar que (16.4.2) se puede ampliar con facilidad a fin de incluir un componente de error aleatorio para tomar en cuenta la variación en el tiempo (véase el ejercicio 16.6).

Los resultados de la estimación del MCE de la función de costo de las aerolíneas se presentan en la tabla 16.5.

Observe estas características del MEFA. El valor (promedio) del intercepto es de 107 429.3. Los valores (diferenciales) del intercepto de las seis entidades se presentan al final de los resultados de la regresión. Por ejemplo, la empresa número 1 tiene un valor de intercepto que es 270 615 unidades menor que el valor del intercepto común de 107 429.3; el valor real del intercepto de esta aerolínea es entonces −163 185.7. Por otro lado, el valor del intercepto de la empresa número 6 es 57 383 unidades mayor que el valor del intercepto común; el valor real del intercepto de esta aerolínea es (107 429.3 + 57 383), o 164 812.3. Los valores de los interceptos de las demás aerolíneas se derivan de modo similar. Sin embargo, debe señalarse que si se suman los valores (diferenciales) de los interceptos de las seis aerolíneas, la suma es 0, como debe ser (¿por qué?).

Si compara los resultados de las regresiones de efectos fijos y efectos aleatorios, observará diferencias considerables entre los dos. La pregunta importante ahora es: ¿qué resultados son confiables? O, de otro modo, ¿qué modelo debe elegirse? Aplicaremos la *prueba de Hausman* para aclarar esta duda.

La hipótesis nula en que se basa la prueba de Hausman es que los estimadores MEF y MCE no difieren considerablemente. El estadístico de prueba desarrollado por Hausman tiene distribu-

[13] Véase Kmenta, *op. cit.*, pp. 625-630.

**TABLA 16.6**

| Efectos aleatorios correlacionados. Prueba de Hausman<br>Ecuación: Sin título<br>Prueba de efectos aleatorios de corte transversal | | | | |
|---|---|---|---|---|
| Resumen de la prueba | | Estadístico ji cuadrada | gl ji cuadrada | Prob. |
| Aleatorios de corte transversal | | 49.619687 | 3 | 0.0000 |
| Comparaciones de la prueba de efectos aleatorios de corte transversal | | | | |
| Variable | Fijos | Aleatorios | Var(dif.) | Prob. |
| *Q* | 3319023.28 | 2288587.95 | 21587779733. | 0.0000 |
| *PF* | 0.773071 | 1.123591 | 0.002532 | 0.0000 |
| *LF* | -3797367.59 | -3084994.0 | 35225469544. | 0.0001 |

ción asintótica $\chi^2$. Si se rechaza la hipótesis nula, la conclusión es que el MCE no es apropiado porque es probable que los efectos aleatorios estén correlacionados con una o más regresoras. En este caso, MEF se prefiere a MCE. En el ejemplo, los resultados de la prueba de Hausman se presentan en la tabla 16.6.

Es claro que la prueba de Hausman rechaza la hipótesis nula, pues el valor estimado de $\chi^2$ para 3 gl es muy significativo; si la hipótesis nula fuera verdadera, la probabilidad de obtener un valor de ji cuadrada de 49.62 o mayor sería prácticamente cero. Como resultado, rechazamos el MCE (MEFA) en favor del MEF. A propósito, la última parte de la tabla precedente compara los coeficientes de efectos fijos y efectos aleatorios de cada variable, y, como indica la última columna, en el presente ejemplo las diferencias son significativas estadísticamente.

### Prueba del multiplicador de Lagrange de Breusch y Pagan[14]

Además de la prueba de Hausman, también podemos aplicar la prueba de Breusch-Pagan (BP) a la hipótesis de que no hay efectos aleatorios, es decir, de que $\sigma_u^2$ en la ecuación (16.6.7) es cero. Los paquetes de software como STATA incorporan esta prueba. Según la hipótesis nula, BP sigue la distribución ji cuadrada con 1 gl; sólo hay 1 gl porque se está probando la hipótesis única de que $\sigma_u^2 = 0$. No presentaremos la fórmula en que se basa la prueba porque es muy complicada.

De nuevo con el ejemplo de las aerolíneas, la aplicación de la prueba BP produce un valor ji cuadrada de 0.61. Con 1 gl, el valor *p* de obtener un valor ji cuadrada de 0.61 o mayor es alrededor de 43%. Por consiguiente, no rechazamos la hipótesis nula. En otras palabras, el modelo de efectos aleatorios no es apropiado en el presente ejemplo. Así, la prueba BP refuerza la prueba de Hausman, con la que también llegamos a la conclusión de que el modelo de efectos aleatorios no es adecuado para el ejemplo de las aerolíneas.

## 16.7 Propiedades de varios estimadores[15]

Analizamos ya varios métodos para estimar modelos de regresión (lineal) de panel, a saber: estimadores agrupados, estimadores de efectos fijos con estimadores de mínimos cuadrados con variable dicótoma (MCVD), estimadores de efectos fijos dentro de grupos, estimadores de primeras diferencias y estimadores de efectos aleatorios. ¿Qué propiedades estadísticas tienen? Como los datos de panel comprenden por lo general una gran cantidad de observaciones, nos concentraremos en la propiedad de consistencia de estos estimadores.

[14] T. Breush y A.R. Pagan, "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics", *Review of Economic Studies,* vol. 47, 1980, pp. 239-253.

[15] El siguiente análisis se basa en A. Colin Cameron y Pravin K. Trivedi, *Microeconometrics: Methods and Applications,* Cambridge University Press, Cambridge, Nueva York, 2005, capítulo 21.

*Estimadores agrupados*

En el supuesto de que los coeficientes de pendiente son constantes a través de todos los sujetos, si el término de error de la ecuación (16.3.1) no está correlacionado con las regresoras, los estimadores agrupados son consistentes. Sin embargo, como ya señalamos, es probable que los términos de error se correlacionen con el paso del tiempo para un sujeto dado. Por consiguiente, deben usarse **errores estándar de panel corregidos** para las pruebas de hipótesis. El lector debe asegurarse de que el paquete estadístico empleado tenga esta facilidad; de lo contrario, pueden subestimarse los errores estándar calculados. Debe subrayarse que si el modelo de efectos fijos es apropiado pero utilizamos el estimador agrupado, los coeficientes estimados serán inconsistentes.

*Estimadores de efectos fijos*

Aunque se suponga que el modelo subyacente es agrupado o aleatorio, los estimadores de efectos fijos siempre son consistentes.

*Estimadores de efectos aleatorios*

El modelo de efectos aleatorios es consistente aunque el verdadero modelo sea el estimador agrupado. Sin embargo, si el verdadero modelo es de efectos fijos, el estimador de efectos aleatorios es inconsistente.

Para demostraciones y mayores detalles sobre estas propiedades, consulte los libros de texto de Cameron y Trivedi, Greene y Wooldridge mencionados en las notas.

## 16.8 Modelo de efectos fijos y modelo de efectos aleatorios: algunos lineamientos

La disyuntiva que enfrenta un investigador es: ¿qué modelo es mejor, MEF o MCE? La respuesta gira en torno del supuesto respecto de la probable correlación entre el componente de error individual, o específico de la unidad de corte transversal, $\varepsilon_i$, y las regresoras $X$.

Si se supone que $\varepsilon_i$ y las $X$ no están correlacionados, el MCE puede resultar apropiado; pero si $\varepsilon_i$ y las $X$ están correlacionados, entonces el MEF puede ser adecuado.

El supuesto en que se basa el MCE es que $\varepsilon_i$ representa una muestra aleatoria de una población mucho más grande, aunque no siempre es así. Por ejemplo, suponga que deseamos estudiar el índice de delincuencia en los 50 estados de Estados Unidos. Como es lógico, en este caso no se sostiene el supuesto de que los 50 estados son una muestra aleatoria.

Teniendo en cuenta esta diferencia fundamental en los dos métodos, ¿qué más podemos decir respecto de la elección entre el MEF y el MCE? A continuación presentamos las observaciones de Judge *et al.*, las cuales pueden resultar de utilidad:[16]

1. Si $T$ (el número de datos de series de tiempo) es grande y $N$ (el número de unidades de corte transversal) es pequeño, es probable que haya muy poca diferencia entre los valores de los parámetros estimados mediante el MEF y el MCE. Por tanto, en este caso la elección se basa en la conveniencia de cálculo. Desde esta perspectiva, parece preferible el MEF.
2. Cuando $N$ es grande y $T$ pequeño (es decir, un panel corto), las estimaciones obtenidas mediante los dos métodos pueden variar de manera significativa. Recuerde que en MCE, $\beta_{1i} = \beta_1 + \varepsilon_i$, donde $\varepsilon_i$ es el componente aleatorio transversal, en tanto que en MEF se considera que $\beta_{1i}$ es fijo y no aleatorio. En el último caso, la inferencia estadística depende de las unidades de corte transversal observadas en la muestra. Lo anterior resulta adecuado si tenemos la firme convicción de que las unidades individuales, o de corte transversal, en la muestra no se extrajeron de manera aleatoria de una muestra mayor. En ese caso, el MEF es adecuado. Sin embargo, si consideramos que las unidades de corte transversal de la muestra se extrajeron de modo aleatorio, el MCE es adecuado, pues aquí la inferencia estadística es incondicional.
3. Si el componente de error individual $\varepsilon_i$ y una o más de las regresoras están correlacionados, los estimadores MCE están sesgados, en tanto que los obtenidos a partir del MEF no lo están.

[16] Judge *et al.*, *op. cit.*, pp. 489-491.

4. Si $N$ es grande y $T$ pequeña, y si los supuestos en los que se basa el MCE son aún válidos, los estimadores MCE son más eficientes que los estimadores MEF.
5. A diferencia del MEF, el MCE puede estimar los coeficientes de variables que no cambian con el tiempo, como el género y el origen étnico. El MEF controla las variables que no cambian con el tiempo, pero no puede estimarlas de manera directa, como evidencian los modelos MCVD y de estimadores dentro de grupos. Por otra parte, MEF controla *todas* las variables invariantes en el tiempo (¿por qué?), mientras que MCE sólo estima las variables invariantes en el tiempo que se introducen explícitamente en el modelo.

Independientemente de la prueba de Hausman, es importante tener en cuenta la advertencia de Johnston y DiNardo. Al decidir entre el modelo de efectos fijos y el de efectos aleatorios, argumentan que "no hay una regla sencilla que ayude al investigador a navegar entre el Escila de los efectos fijos y el Caribdis del error de medición y la selección dinámica. Sin embargo, aunque representan una mejora respecto de los datos de corte transversal, los datos de panel no son una cura milagrosa para todos los problemas de los econometristas".[17]

## 16.9 Regresiones con datos de panel: algunos comentarios para concluir

Como mencionamos al principio, el tema de la creación de modelos con datos de panel es vasto y complejo. Apenas tocamos la superficie. Entre los puntos que no hemos analizado destacan los siguientes:

1. Pruebas de hipótesis con datos de panel.
2. Heteroscedasticidad y autocorrelación en MCE.
3. Datos de panel desbalanceado.
4. Modelos dinámicos con datos de panel en los que el (los) valor(es) rezagado(s) de la regresada aparecen como una variable explicativa.
5. Ecuaciones simultáneas que se relacionan con datos de panel.
6. Variables dependientes cualitativas y datos de panel.
7. Raíces unitarias en datos de panel (en relación con las raíces unitarias, véase el capítulo 21).

Uno o más de estos temas se encuentran en las referencias citadas en este capítulo, por lo que se exhorta al lector a que las consulte a fin de aprender más sobre dichos temas. Las referencias también citan diversos estudios empíricos en diversas áreas de negocios y económicas mediante modelos de regresión con datos de panel. Se aconseja al principiante que lea algunas de esas aplicaciones con el fin de darse una idea de la forma en que los investigadores han puesto en práctica dichos modelos.[18]

## 16.10 Algunos ejemplos ilustrativos

**EJEMPLO 16.1**
*Productividad e inversión pública*

Para averiguar por qué disminuye la productividad y qué función desempeña la inversión pública, Alicia Munnell estudió datos de productividad en los 48 estados del territorio continental de Estados Unidos durante 17 años, de 1970 a 1986, para un total de 816 observaciones.[19] Con estos datos estimamos la regresión agrupada de la tabla 16.7. Observe que esta regresión no toma en cuenta la naturaleza de panel de los datos.

La variable dependiente de este modelo es PEB (producto estatal bruto), y las variables explicativas son CAPPRIV (capital privado), CAPPUB (capital público), AGUA (capital del servicio de agua) y DESEMP (tasa de desempleo). *Nota: L* representa el logaritmo natural.

(*continúa*)

[17] Jack Johnston y John DiNardo, *Econometric Methods*, 4a. ed., McGraw-Hill, 1997, p. 403.

[18] Para más detalles y aplicaciones concretas, véase Paul D. Allison, *Fixed Effects Regression Methods for Longitudinal Data, Using SAS,* SAS Institute, Cary, Carolina del Norte, 2005.

[19] Los datos de Munnell se encuentran en www.aw-bc.com/murray.

**EJEMPLO 16.1** (*continuación*)

## TABLA 16.7

Variable dependiente: LPEB
Método: Mínimos cuadrados de panel
Muestra: 1970-1986
Periodos incluidos: 17
Secciones de corte transversal incluidas: 48
Total de observaciones de panel (balanceado): 816

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | 0.907604 | 0.091328 | 9.937854 | 0.0000 |
| *L*CAPPRIV | 0.376011 | 0.027753 | 13.54847 | 0.0000 |
| *L*CAPPUB | 0.351478 | 0.016162 | 21.74758 | 0.0000 |
| *L*AGUA | 0.312959 | 0.018739 | 16.70062 | 0.0000 |
| *L*DESEMP | -0.069886 | 0.015092 | -4.630528 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.981624 | Media de la variable dependiente | 10.50885 |
| *R* cuadrada ajustada | 0.981533 | Desviación estándar de la variable dependiente | 1.021132 |
| Error estándar de la regresión | 0.138765 | Estadístico *F* | 10830.51 |
| Suma de cuadrados residual | 15.61630 | Prob. (estadístico *F*) | 0.000000 |
| Log verosimilitud | 456.2346 | Estadístico Durbin-Watson | 0.063016 |

Todas las variables tienen los signos esperados y todas son estadísticamente significativas, tanto en lo individual como colectivamente, si se asumen válidos todos los supuestos del modelo clásico de regresión lineal.

Para tomar en cuenta la dimensión de panel de los datos, en la tabla 16.8 se estimó un modelo de efectos fijos con 47 variables dicótomas para los 48 estados a fin de evitar caer en

## TABLA 16.8

Variable dependiente: LPEB
Método: Mínimos cuadrados de panel
Muestra: 1970-1986
Periodos incluidos: 17
Unidades de corte transversal incluidas: 48
Total de observaciones de panel (balanceado): 816

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | -0.033235 | 0.208648 | -0.159286 | 0.8735 |
| *L*CAPPRIV | 0.267096 | 0.037015 | 7.215864 | 0.0000 |
| *L*CAPPUB | 0.714094 | 0.026520 | 26.92636 | 0.0000 |
| *L*AGUA | 0.088272 | 0.021581 | 4.090291 | 0.0000 |
| *L*DESEMP | -0.138854 | 0.007851 | -17.68611 | 0.0000 |

Especificación de efectos

Fijos de corte transversal (variables dicótomas)

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.997634 | Media de la variable dependiente | 10.50885 |
| *R* cuadrada ajustada | 0.997476 | Desviación estándar de la variable dependiente | 1.021132 |
| Error estándar de la regresión | 0.051303 | Estadístico *F* | 6315.897 |
| Suma de cuadrados residual | 2.010854 | Prob. (estadístico *F*) | 0.000000 |
| Log verosimilitud | 1292.535 | Estadístico Durbin-Watson | 0.520682 |

**EJEMPLO 16.1**
(*continuación*)

**TABLA 16.9**

Variable dependiente: LPEB
Método: Panel EMCG (efectos aleatorios de corte transversal)
Muestra: 1970-1986
Periodos incluidos: 17
Unidades de corte transversal incluidas: 48
Total de observaciones de panel (balanceado): 816
Estimador Swamy y Arora de varianzas de componentes

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | -0.046176 | 0.161637 | -0.285680 | 0.7752 |
| *L*CAPPRIV | 0.313980 | 0.029740 | 10.55760 | 0.0000 |
| *L*CAPPUB | 0.641926 | 0.023330 | 27.51514 | 0.0000 |
| *L*AGUA | 0.130768 | 0.020281 | 6.447875 | 0.0000 |
| *L*DESEMP | -0.139820 | 0.007442 | -18.78669 | 0.0000 |

Especificación de efectos

| | D.E | Rho |
|---|---|---|
| Aleatorios de corte transversal | 0.130128 | 0.8655 |
| Aleatorios idiosincrásicos | 0.051303 | 0.1345 |

la trampa de las variables dicótomas. Para ahorrar espacio, sólo presentamos los coeficientes estimados de la regresión y no los coeficientes de cada variable dicótoma. Sin embargo, debe añadirse que las 47 variables dicótomas estatales fueron muy significativas estadísticamente en lo individual.

Se observan diferencias considerables entre la regresión agrupada y la regresión de efectos fijos, lo que despierta dudas sobre los resultados de la regresión agrupada.

Para ver si el modelo de efectos aleatorios es más apropiado en este caso, en la tabla 16.9 presentamos los resultados del modelo de regresión de efectos aleatorios.

Para elegir entre los dos modelos aplicamos la prueba de Hausman, que arrojó los resultados de la tabla 16.10.

Como el valor estimado ji cuadrada es muy significativo estadísticamente, rechazamos la hipótesis de que no hay diferencia significativa en los coeficientes estimados de los dos modelos. Al parecer, existe correlación entre el término de error y una o más de las regresoras. Por tanto, podemos rechazar el modelo de efectos aleatorios y elegir el modelo de efectos fijos. Sin embargo, hay que tener en cuenta, como muestra la última parte de la tabla 16.10, que no todos los coeficientes difieren en los dos modelos. Por ejemplo, no existe diferencia significativa estadísticamente en los valores del coeficiente de *L*DESEMP en los dos modelos.

**TABLA 16.10**

| Resumen de la prueba | Estadístico ji cuadrada | gl ji cuadrada | Prob. |
|---|---|---|---|
| Aleatorios de corte transversal | 42.458353 | 4 | 0.0000 |

Comparaciones de la prueba de efectos aleatorios de corte transversal:

| Variable | Fijos | Aleatorios | Var(dif.) | Prob. |
|---|---|---|---|---|
| *L*CAPPRIV | 0.267096 | 0.313980 | 0.000486 | 0.0334 |
| *L*CAPPUB | 0.714094 | 0.641926 | 0.000159 | 0.0000 |
| *L*AGUA | 0.088272 | 0.130768 | 0.000054 | 0.0000 |
| *L*DESEMP | −0.138854 | −0.139820 | 0.000006 | 0.6993 |

**EJEMPLO 16.2**
*Demanda de electricidad en Estados Unidos*

En su artículo, Maddala *et al.* consideraron la demanda residencial de electricidad y gas natural en 49 estados de Estados Unidos de 1970 a 1990; Hawai no se incluyó en el análisis.[20] Recopilaron datos sobre diversas variables; estos datos se encuentran en el sitio web del libro. En este ejemplo sólo consideraremos la demanda residencial de electricidad. Primero presentamos los resultados con base en la estimación de efectos fijos (tabla 16.11) y después la estimación de efectos aleatorios (tabla 16.12), seguidas de una comparación de los dos modelos.

**TABLA 16.11**

Variable dependiente: Log(CREPC)
Método: Mínimos cuadrados de panel

Muestra: 1971-1990
Periodos incluidos: 20
Unidades de corte transversal incluidas: 49
Total de observaciones de panel (balanceado): 980

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | -12.55760 | 0.363436 | -34.55249 | 0.0000 |
| Log(PRE) | -0.628967 | 0.029089 | -21.62236 | 0.0000 |
| Log(IDRPC) | 1.062439 | 0.040280 | 26.37663 | 0.0000 |

Especificación de efectos

Fijos de corte transversal (variables dicótomas)

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.757600 | Media de la variable dependiente | -4.536187 |
| *R* cuadrada ajustada | 0.744553 | Desviación estándar de la variable dependiente | 0.316205 |
| Error estándar de la regresión | 0.159816 | Criterio de información de Akaike | -0.778954 |
| Suma de cuadrados residual | 23.72762 | Criterio de Schwarz | -0.524602 |
| Log verosimilitud | 432.6876 | Criterio de Hannan-Quinn | -0.682188 |
| Estadístico *F* | 58.07007 | Estadístico Durbin-Watson | 0.404314 |
| Prob. (estadístico *F*) | 0.000000 | | |

donde Log(CREPC) = logaritmo natural del consumo residencial de electricidad *per cápita* (millones de btu = unidades térmicas británicas), Log(PRE) = logaritmo natural del precio real de la electricidad en 1987 y Log(IDRPC) = logaritmo natural del ingreso disponible real *per cápita* en 1987.

Como se trata de un modelo de doble logaritmo, los coeficientes estimados de pendiente representan elasticidades. Por tanto, cuando todos los demás factores permanecen constantes, si el ingreso real *per cápita* aumenta 1%, la media del consumo de electricidad aumenta casi 1%. Asimismo, cuando todo lo demás permanece constante, si el precio real de la electricidad aumenta 1%, el consumo promedio de electricidad disminuye alrededor de 0.6 por ciento. Todas las elasticidades estimadas son significativas estadísticamente.

Los resultados del modelo de error aleatorio se presentan en la tabla 16.12.

Al parecer, no hay mucha diferencia entre los dos modelos. Pero con la prueba de Hausman podemos averiguar si esto es verdad. Los resultados de la prueba aparecen en la tabla 16.13.

Aunque los coeficientes de los dos modelos de las tablas 16.11 y 16.12 dan la impresión de ser muy parecidos, la prueba de Hausman demuestra que no es así. El valor ji cuadrada es muy significativo estadísticamente. Por tanto, elegimos el modelo de efectos fijos y no el de efectos

[20] G.S. Maddala, Robert P. Trost, Hongyi Li y Frederick Joutz, "Estimation of Short-run and Long-run Elasticities of Demand from Panel Data Using Shrikdage Estimators", *Journal of Business and Economic Statistics,* vol. 15, núm. 1, enero de 1997, pp. 90-100.

**EJEMPLO 16.2**
(*continuación*)

**TABLA 16.12**

Variable dependiente: Log(CREPC)
Método: Panel EMCG (efectos aleatorios de corte transversal)
Muestra: 1971-1990
Periodos incluidos: 20
Unidades de corte transversal incluidas: 49
Total de observaciones de panel (balanceado): 980
Estimador Swamy y Arora de varianzas de componentes

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | -11.68536 | 0.353285 | -33.07631 | 0.0000 |
| Log(PRE) | -0.665570 | 0.028088 | -23.69612 | 0.0000 |
| Log(IDRPC) | 0.980877 | 0.039257 | 24.98617 | 0.0000 |

Especificación de efectos

| | D.E | Rho |
|---|---|---|
| Aleatorios de corte transversal | 0.123560 | 0.3741 |
| Aleatorios idiosincrásicos | 0.159816 | 0.6259 |

Estadísticas ponderadas

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.462591 | Media de la variable dependiente | -1.260296 |
| *R* cuadrada ajustada | 0.461491 | Desviación estándar de la variable dependiente | 0.229066 |
| Error estándar de la regresión | 0.168096 | Suma de cuadrados residual | 27.60641 |
| Estadístico *F* | 420.4906 | Estadístico Durbin-Watson | 0.345453 |
| Prob. (estadístico *F*) | 0.000000 | | |

Estadísticas sin ponderar

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.267681 | Media de la variable dependiente | -4.536187 |
| Suma de cuadrados residual | 71.68384 | Estadístico Durbin-Watson | 0.133039 |

**TABLA 16.13**

Efectos aleatorios correlacionados. Prueba de Hausman
Ecuación: Sin título
Prueba de efectos aleatorios de corte transversal

| Resumen de la prueba | Estadístico ji cuadrada | gl ji cuadrada | Prob. |
|---|---|---|---|
| Aleatorios de corte transversal | 105.865216 | 2 | 0.0000 |

Comparaciones de la prueba de efectos aleatorios de corte transversal:

| Variable | Fijos | Aleatorios | Var(dif.) | Prob. |
|---|---|---|---|---|
| Log(PRE) | -0.628967 | -0.665570 | 0.000057 | 0.0000 |
| Log(IDRPC) | 1.062439 | 0.980877 | 0.000081 | 0.0000 |

aleatorios. *Este ejemplo destaca el punto importante de que cuando el tamaño de la muestra es grande, en este caso 980 observaciones, hasta las diferencias más pequeñas en los coeficientes estimados de los dos modelos pueden ser estadísticamente significativas.* En consecuencia, los coeficientes de la variable Log(PRE) de los dos modelos se ven razonablemente parecidos, pero estadísticamente no lo son.

**EJEMPLO 16.3**
*Consumo de cerveza, ingreso e impuesto sobre ventas de cerveza*

Para evaluar el efecto del impuesto sobre ventas de cerveza en el consumo de esta bebida, Philip Cook investigó la relación entre los dos después de tener en cuenta el efecto del ingreso.[21] Los datos corresponden a los 50 estados de Estados Unidos y Washington, D.C., de 1975 a 2000. En este ejemplo estudiamos la relación de las ventas de cerveza *per cápita* y la tasa impositiva y el ingreso, todo en el ámbito estatal. Se presentan los resultados de los modelos de MCO agrupados, efectos fijos y efectos aleatorios en forma tabular en la tabla 16.14. La variable dependiente es ventas de cerveza *per cápita*.

Estos resultados son interesantes. De acuerdo con la teoría económica, se espera una relación negativa entre el consumo de cerveza y los impuestos sobre ventas de cerveza, que es precisamente lo que ocurre en los tres modelos. El efecto negativo del ingreso sobre el consumo de cerveza indica que la cerveza es un **bien inferior**, es decir, un producto cuya demanda decrece a medida que el ingreso del consumidor aumenta. Tal vez cuando el ingreso aumenta, los consumidores prefieren champaña.

Para los efectos que aquí interesan, lo importante es la diferencia entre los coeficientes estimados. En apariencia, no hay mucha diferencia entre los coeficientes estimados con el MEF y MCE. De hecho, la prueba de Hausman produce un valor ji cuadrada de 3.4, que no es significativo para 2 gl en el nivel de 5%; el valor $p$ es 0.1783.

Sin embargo, los resultados basados en MCO son muy diferentes. El coeficiente de la variable impuesto sobre ventas de cerveza, en valor absoluto, es mucho más pequeño del que se obtiene con el MEF o con el MCE. La variable ingreso, aunque tiene signo negativo, no es estadísticamente significativa, mientras que los otros dos modelos muestran que es muy significativa.

Este ejemplo revela de manera muy vívida lo que puede ocurrir si descuidamos la estructura de panel de los datos y estimamos una regresión agrupada.

**TABLA 16.14**

| Variable | MCO | MEF | MEFA |
|---|---|---|---|
| Constante | 1.4192 | 1.7617 | 1.7542 |
| | (24.37) | (52.23) | (39.22) |
| Impuesto s/ventas cerveza | −0.0067 | −0.0183 | −0.0181 |
| | (−2.13) | (−9.67) | (−9.69) |
| Ingreso | $-3.54(e^{-6})$ | −0.000020 | −0.000019 |
| | (−1.12) | (−9.17) | (−9.10) |
| $R^2$ | 0.0062 | 0.0052 | 0.0052 |

*Notas:* Las cifras entre paréntesis son las razones $t$ estimadas. $-3.54(e^{-6}) = -0.00000354$.

## Resumen y conclusiones

1. Los modelos de regresión de panel se basan en los datos de panel, los cuales consisten en observaciones sobre las mismas unidades de corte transversal, o individuales, a lo largo de varios periodos.
2. Existen diversas ventajas en los datos de panel. Primera, incrementan de modo considerable el tamaño de la muestra. Segunda, al estudiar observaciones de corte transversal repetidas, los datos de panel resultan más adecuados para estudiar las dinámicas del cambio. Tercera, los datos de panel permiten estudiar modelos de comportamiento más complejos.
3. A pesar de sus ventajas sustanciales, los datos de panel plantean diversos problemas de estimación y de inferencia. Como esos datos implican dimensiones de corte transversal y temporales, necesitan abordarse los problemas que plagan a los datos de corte transversal (por ejemplo, la heteroscedasticidad) y los datos de series de tiempo (por ejemplo, la autocorrelación). Además hay otros problemas, como la correlación cruzada en unidades individuales en el mismo punto en el tiempo.

[21] Estos datos provienen del sitio web de Michael P. Murphy, *Econometrics: A Modern Introduction,* Pearson/Addison Wesley, Boston, 2006, pero los datos originales fueron recopilados por Philip Cook para su libro *Paying the Tab: The Costs and Benefits of Alcohol Control,* Princeton University Press, Princeton, Nueva Jersey, 2007.

4. Hay varias técnicas de estimación para abordar uno o más de estos problemas. Las dos más importantes son: 1) el modelo de efectos fijos (MEF) y 2) el modelo de efectos aleatorios (MEFA) o modelo de componentes del error (MCE).
5. En el MEF se permite que el intercepto en el modelo de regresión difiera entre individuos, a manera de reconocimiento de que cada unidad individual, o transversal, pueda tener algunas características especiales por sí mismas. A fin de tomar en cuenta los distintos interceptos, se pueden utilizar variables dicótomas. El MEF que emplea esas variables se conoce como modelo de mínimos cuadrados con variable dicótoma (MCVD). El MEF resulta apropiado en situaciones donde el intercepto específico individual puede estar correlacionado con una o más regresoras. Una desventaja del modelo MCVD es que consume muchos grados de libertad cuando el número de unidades de corte transversal, N, es muy grande, en cuyo caso se tendrán que introducir N variables dicótomas (pero habrá que suprimir el término del intercepto común).
6. Una alternativa al MEF es el MCE. En este último, se supone que el intercepto de una unidad individual se extrae de manera aleatoria de una población mucho más grande con un valor medio constante. Así, el intercepto individual se expresa como una desviación respecto de este valor medio constante. Una ventaja del MCE respecto del MEF consiste en la economía de los grados de libertad, en vista de que no se tiene que calcular N interceptos de corte transversal. Sólo se requiere estimar el valor medio del intercepto y su varianza. El MCE es adecuado para situaciones en las que el intercepto (aleatorio) de cada unidad de corte transversal no está correlacionado con las regresoras. Otra ventaja del MCE es que se pueden introducir variables, como género, religión y origen étnico, que se mantienen constantes en cualquier sujeto dado. En el MEF no podemos hacer eso, porque todas esas variables son colineales con el intercepto específico del sujeto. Además, si usamos el estimador dentro de grupos o el estimador de primeras diferencias, se borrará por completo toda esa invariabilidad con el tiempo.
7. Con la prueba de Hausman se decide entre el MEF y el MCE. También se puede aplicar la prueba de Breusch-Pagan para ver si el MCE es apropiado.
8. A pesar de su creciente popularidad en la investigación aplicada, y no obstante la creciente disponibilidad de esos datos, las regresiones con datos de panel tal vez no sean adecuadas para todas las situaciones. Se tiene que echar mano de un criterio práctico en cada caso.
9. Existen algunos problemas específicos con los datos de panel que deben tenerse presentes. El más grave es el problema de desgaste, en el que, por una razón u otra, los sujetos del panel se retiran a medida que pasa el tiempo, de modo que en las encuestas subsiguientes (o de corte transversal) hay menos sujetos originales en el panel. Aunque no haya desgaste, con el transcurso del tiempo los sujetos pueden negarse o no estar dispuestos a responder algunas preguntas.

## EJERCICIOS

### Preguntas

16.1. ¿Cuáles son las características especiales de *a*) los datos de corte transversal, *b*) las series de tiempo y *c*) los datos de panel?

16.2. ¿Qué se quiere dar a entender con modelo de efectos fijos (MEF)? Como los datos de panel presentan las dos dimensiones de tiempo y espacio, ¿cómo es que el MEF permite ambas dimensiones?

16.3. ¿Qué se quiere dar a entender con modelo con componentes del error (MCE)? ¿En qué difiere del MEF? ¿Cuándo resulta apropiado el MCE, y cuándo el MEF?

16.4. ¿Hay diferencia entre los modelos de mínimos cuadrados con variables dicótomas (MCVD), estimados dentro de grupos y primeras diferencias?

16.5. ¿Cuándo resultan inapropiados los modelos de regresión con datos de panel? Proporcione ejemplos.

16.6. ¿Cómo ampliaría el modelo (16.4.2) para incluir un componente del error en el tiempo? Escriba el modelo explícitamente.

16.7. Consulte el ejemplo de los huevos y sus precios de la tabla 1.1. ¿Qué modelo resultaría apropiado en este caso, el MEF o el MCE? ¿Por qué?

16.8. Para los datos de inversión presentados en la tabla 1.2, ¿qué modelo debe elegirse, MEF o MEFA? ¿Por qué?

16.9. Con base en el Michigan Income Dynamics Study (Estudio Michigan de dinámica del ingreso), Hausman trató de estimar un modelo para salarios, o ganancias, con una muestra de 629 egresados del nivel medio superior, a quienes se les dio un seguimiento durante seis años, lo cual dio como resultado un total de 3 774 observaciones. En este estudio, la variable dependiente fue el logaritmo del salario y las variables explicativas fueron edad (dividida en varios grupos de edad), desempleo en el año anterior, pobreza sanitaria en el año anterior, autoempleo, región de residencia (Sur = 1; 0 en otro caso), área de residencia (rural = 1; 0 en otro caso). Hausman utilizó el MEF y el MCE. Estos resultados se proporcionan en la tabla 16.15 (se dan los errores estándar entre paréntesis).

**TABLA 16.15**
**Ecuaciones de salarios (variable dependiente: log salario)**

Fuente: Reproducido de Cheng Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986, p. 42. Fuente original: J. A. Hausman, "Specification Tests in Econometrics", *Econometrica*, vol. 46, 1978, pp. 1251-1271.

| | Variable | Efectos fijos | Efectos aleatorios |
|---|---|---|---|
| 1. | Edad 1 (20-35) | 0.0557 (0.0042) | 0.0393 (0.0033) |
| 2. | Edad 2 (35-45) | 0.0351 (0.0051) | 0.0092 (0.0036) |
| 3. | Edad 3 (45-55) | 0.0209 (0.0055) | −0.0007 (0.0042) |
| 4. | Edad 4 (55-65) | 0.0209 (0.0078) | −0.0097 (0.0060) |
| 5. | Edad 5 (65– ) | −0.0171 (0.0155) | −0.0423 (0.0121) |
| 6. | Desempleo en el año anterior | −0.0042 (0.0153) | −0.0277 (0.0151) |
| 7. | Pobreza sanitaria en el año anterior | −0.0204 (0.0221) | −0.0250 (0.0215) |
| 8. | Autoempleo | −0.2190 (0.0297) | −0.2670 (0.0263) |
| 9. | Sur | −0.1569 (0.0656) | −0.0324 (0.0333) |
| 10. | Rural | −0.0101 (0.0317) | −0.1215 (0.0237) |
| 11. | Constante | — — | 0.8499 (0.0433) |
| | $S^2$ | 0.0567 | 0.0694 |
| | Grados de libertad | 3 135 | 3 763 |

*a*) ¿Los resultados tienen sentido económico?

*b*) ¿Existe una gran diferencia en los resultados producidos por los dos modelos? Si así fuera, ¿qué explicaría tales diferencias?

*c*) Con base en los datos de la tabla, ¿qué modelo, si acaso existiera uno, elegiría?

## Ejercicios empíricos

16.10. Consulte el ejemplo de las aerolíneas analizado en el texto. En lugar del modelo lineal presentado en la ecuación (16.4.2), estime un modelo de regresión log-lineal y compare los resultados que se proporcionan en la tabla 16.2

16.11. Consulte los datos de la tabla 1.1.

*a*) Sea $Y$ = huevos producidos (en millones) y $X$ = precio de los huevos (centavos por docena). Calcule el modelo para los años 1990 y 1991 por separado.

*b*) Agrupe las observaciones para los dos años y estime la regresión agrupada. ¿Qué suposiciones hace al agrupar los datos?

*c*) Utilice el modelo de efectos fijos con la distinción entre los dos años y presente los resultados de la regresión.

*d*) ¿Puede utilizar el modelo de efectos fijos si hace la distinción de los 50 estados? ¿Por qué?

*e*) ¿Tiene sentido distinguir el efecto por estado y el efecto por año? De ser así, ¿cuántas variables dicótomas tendría que introducir?

*f*) ¿El modelo de componentes del error sería apropiado como modelo para la producción de huevos? ¿Por qué? Vea si puede estimar dicho modelo con *EViews*, por ejemplo.

16.12. Continúe con el ejercicio 16.11. Antes de decidir hacer la regresión agrupada, quizá desee averiguar si los datos son "agrupables". Para este propósito, se decide por la prueba de Chow analizada en el capítulo 8. Muestre los cálculos necesarios y determine si la regresión agrupada tiene sentido.

16.13. Use los datos de inversión de la tabla 1.6.

*a*) Estime la función de inversión Grunfeld para cada empresa de manera individual.

*b*) Ahora agrupe los datos de todas las empresas y estime la función de inversión Grunfeld por MCO.

*c*) Estime la función de inversión con MCVD y compare los resultados con la regresión agrupada que estimó en *b*).

*d*) ¿Cómo decidiría entre la regresión agrupada y la regresión de MCVD? Realice los cálculos necesarios.

16.14. La tabla 16.16 suministra datos sobre la tasa de desempleo civil $Y$ (%) y los salarios por hora del sector manufacturero en dólares estadounidenses $X$ (índice, 1992 = 100) para Canadá, Reino Unido y Estados Unidos de 1980 a 2006. Considere el modelo:

$$Y_{it} = \beta_1 + \beta_2 X_{it} + u_{it} \quad (1)$$

**TABLA 16.16**
**Tasa de desempleo y salario por hora en el sector manufacturero, en Estados Unidos, Canadá y Reino Unido, 1980-2006.**

Fuente: *Economic Report of the President*, enero de 2008, tabla B-109.

| Año | Sal_EUA | Desemp_EUA | Sal_CAN | Desemp_CAN | Sal_RU | Desemp_RU |
|---|---|---|---|---|---|---|
| 1980 | 55.9 | 7.1 | 49.0 | 7.3 | 47.1 | 6.9 |
| 1981 | 61.6 | 7.6 | 53.8 | 7.3 | 47.5 | 9.7 |
| 1982 | 67.2 | 9.7 | 60.1 | 10.7 | 45.1 | 10.8 |
| 1983 | 69.3 | 9.6 | 64.3 | 11.6 | 41.9 | 11.5 |
| 1984 | 71.6 | 7.5 | 65.0 | 10.9 | 39.8 | 11.8 |
| 1985 | 75.3 | 7.2 | 65.0 | 10.2 | 42.3 | 11.4 |
| 1986 | 78.8 | 7.0 | 64.9 | 9.3 | 52.0 | 11.4 |
| 1987 | 81.3 | 6.2 | 69.6 | 8.4 | 64.5 | 10.5 |
| 1988 | 84.1 | 5.5 | 78.5 | 7.4 | 74.8 | 8.6 |
| 1989 | 86.6 | 5.3 | 85.5 | 7.1 | 73.5 | 7.3 |
| | | 5.6 | | | | |
| 1990 | 90.5 | | 92.4 | 7.7 | 89.6 | 7.1 |
| 1991 | 95.6 | 6.8 | 100.7 | 9.8 | 99.9 | 8.9 |
| 1992 | 100.0 | 7.5 | 100.0 | 10.6 | 100.0 | 10.0 |
| 1993 | 102.0 | 6.9 | 94.8 | 10.8 | 88.8 | 10.4 |
| | | 6.1 | | | | |
| 1994 | 105.3 | | 92.1 | 9.6 | 92.8 | 8.7 |
| 1995 | 107.3 | 5.6 | 93.9 | 8.6 | 97.3 | 8.7 |
| 1996 | 109.3 | 5.4 | 95.9 | 8.8 | 96.0 | 8.1 |
| 1997 | 112.2 | 4.9 | 96.7 | 8.4 | 104.1 | 7.0 |
| 1998 | 118.7 | 4.5 | 94.9 | 7.7 | 113.8 | 6.3 |
| 1999 | 123.4 | 4.2 | 96.8 | 7.0 | 117.5 | 6.0 |
| 2000 | 134.7 | 4.0 | 100.0 | 6.1 | 114.8 | 5.5 |
| 2001 | 137.8 | 4.7 | 98.9 | 6.5 | 114.7 | 5.1 |
| 2002 | 147.8 | 5.8 | 101.0 | 7.0 | 126.8 | 5.2 |
| 2003 | 158.2 | 6.0 | 116.7 | 6.9 | 145.2 | 5.0 |
| 2004 | 161.5 | 5.5 | 127.1 | 6.4 | 171.4 | 4.8 |
| 2005 | 168.3 | 5.1 | 141.8 | 6.0 | 177.4 | 4.8 |
| 2006 | 172.4 | 4.6 | 155.5 | 5.5 | 192.3 | 5.5 |

*Notas:* Desemp = Tasa de desempleo %. Sal = Índice de salarios por hora en dólares estadounidenses, 1992-100. CAN = Canadá. RU = Reino Unido.

*a*) *A priori*, ¿cuál es la relación esperada entre $Y$ y $X$? ¿Por qué?

*b*) Calcule el modelo dado en (1) para cada país.

*c*) Estime el modelo agrupando las 81 observaciones en su totalidad.

*d*) Calcule el modelo de efectos fijos.

*e*) Estime el modelo de componentes del error.

*f*) ¿Cuál es el mejor modelo, el MEF o el MCE? Justifique la respuesta. (*Sugerencia:* Aplique la prueba de Hausman.)

16.15. Baltagi y Griffin consideraron la siguiente función de demanda de gasolina:*

$$\ln Y_{it} = \beta_1 + \beta_2 \ln X_{2it} + \beta_3 \ln X_{3it} + \beta_4 \ln X_{4it} + u_{it}$$

donde $Y =$ consumo de gasolina por automóvil; $X_2 =$ ingreso real *per cápita*, $X_3 =$ precio real de la gasolina, $X_4 =$ número de automóviles *per cápita*, $i =$ clave del país, en los 18 países miembros de la OCDE, y $t =$ tiempo (observaciones anuales de 1960 a 1978). *Nota:* Los valores de la tabla ya muestran el rezago correspondiente.

*a*) Estime la función de demanda anterior agrupando los datos de los 18 países miembros (un total de 342 observaciones).

*b*) Estime el modelo de efectos fijos con los mismos datos.

*c*) Estime el modelo de componentes aleatorios con los mismos datos.

*d*) Con base en este análisis, ¿qué modelo describe mejor la demanda de gasolina en los 18 países de la OCDE? Justifique su respuesta.

16.16. El artículo de Subhayu Bandyopadhyay y Howard J. Wall "The Determinants of Aid in the Post-Cold War Era", *Review,* Federal Reserve Bank of St. Louis, noviembre-diciembre de 2007, vol. 89, número 6, pp. 533-547, estima, con datos de panel, la capacidad de ayuda a las necesidades económicas y físicas de los países beneficiarios, los derechos civiles y políticos, y la eficacia del gobierno. Los datos corresponden a 135 países y abarcan tres años. El artículo y los datos se encuentran en: http://research.stlouisfed.org/publications/review/past/2007 en la sección núm 10, vol. 89, noviembre y diciembre. Los datos también se encuentran en el sitio web de este libro, en la tabla 16.18. Estime el modelo de los autores (presentado en la página 534 del artículo) mediante un estimador de efectos aleatorios. Compare los resultados con los de los estimadores de efectos agrupados y fijos proporcionados por los autores en la tabla 2 de su artículo. ¿Qué modelo es apropiado en este caso, el de efectos fijos o el de efectos aleatorios? ¿Por qué?

16.17. Consulte el ejemplo de las aerolíneas analizado en el texto. Para cada aerolínea, estime una función de costo logarítmica de series de tiempo. ¿Cómo se comparan estas regresiones con los modelos de efectos fijos y efectos aleatorios estudiados en el capítulo? ¿También puede estimar 15 funciones logarítmicas de costo de corte transversal ? ¿Por qué?

* B.H. Baltagi y J.M. Griffin, "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures", *European Economic Review*, vol. 22, 1983, pp. 117-137. Los datos de los 18 países de la OCDE correspondientes a los años 1960-1978 se encuentran en http://www.wiley.com/legacy/wileychi/baltagi/supp/Gasoline.dat, o en el sitio web de este texto, tabla 16.17.

Capítulo 17

# Modelos econométricos dinámicos: modelos autorregresivos y de rezagos distribuidos

En el análisis de regresión con datos de series de tiempo, cuando el modelo de regresión incluye no sólo valores actuales sino además valores rezagados (pasados) de las variables explicativas (las *X*), se denomina **modelo de rezagos distribuidos**. Si el modelo incluye uno o más valores rezagados de la variable dependiente entre sus variables explicativas, se denomina **modelo autorregresivo**. Así,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$$

representa un modelo de rezagos distribuidos, mientras que

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t$$

es un ejemplo de modelo autorregresivo. Estos últimos también se conocen como **modelos dinámicos**, pues señalan la trayectoria en el tiempo de la variable dependiente en relación con su(s) valor(es) pasado(s).

Los modelos autorregresivos y de rezagos distribuidos son muy comunes en el análisis econométrico, y en este capítulo los estudiaremos en detalle con el objeto de averiguar lo siguiente:

1. ¿Cuál es el papel de los rezagos en economía?
2. ¿Con qué razones se justifican los rezagos?
3. ¿Existe alguna justificación teórica para los modelos rezagados comunes en la econometría empírica?
4. ¿Cuál es la relación, si acaso, entre los modelos autorregresivos y los modelos de rezagos distribuidos? ¿Pueden derivarse unos de otros?
5. ¿Cuáles son algunos problemas estadísticos relacionados con la estimación de tales modelos?
6. ¿La relación adelantada-rezagada entre variables implica causalidad? De ser así, ¿cómo se mide?

## 17.1 El papel del "tiempo" o "rezago" en economía

En economía, la dependencia de una variable $Y$ (la variable dependiente) respecto de otra u otras variables $X$ (las variables explicativas) pocas veces es instantánea. Con frecuencia $Y$ responde a $X$ en un lapso, el cual se denomina rezago. Para ilustrar la naturaleza del rezago, aquí consideramos diversos ejemplos.

**EJEMPLO 17.1**
*La función consumo*

Suponga que una persona recibe un incremento salarial de $2 000 en su pago anual y que se trata de un incremento "permanente" en el sentido de que se mantiene el incremento en el salario. ¿Cuál será el efecto de este incremento en el ingreso sobre su gasto de consumo anual?

Después del aumento en el ingreso, la gente no se apura a gastarse todo el incremento de inmediato. Así, el beneficiario de este ejemplo puede decidir aumentar su gasto de consumo $800 durante el primer año después del incremento en el ingreso, $600 en el siguiente año y otros $400 un año después, para ahorrar el resto. A finales del tercer año, el gasto de consumo anual de la persona habrá aumentado $1 800. Entonces la función de consumo se escribe

$$Y_t = \text{constante} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + u_t \qquad \textbf{(17.1.1)}$$

donde $Y$ es el gasto de consumo y $X$ es el ingreso.

La ecuación (17.1.1) muestra que el efecto de un incremento de $2 000 en el ingreso se propaga, o distribuye, durante un periodo de tres años. Por consiguiente, modelos como (17.1.1) se denominan **modelos de rezagos distribuidos**, porque el efecto de una causa dada (el ingreso) se propaga durante varios periodos. En términos geométricos, el modelo de rezagos distribuidos (17.1.1) se muestra en la figura 17.1, o, alternativamente, en la figura 17.2.

**FIGURA 17.1**
Ejemplo de rezagos distribuidos.

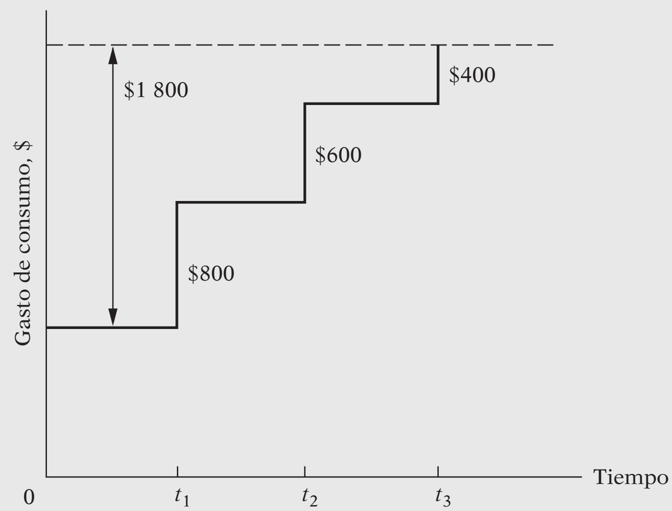
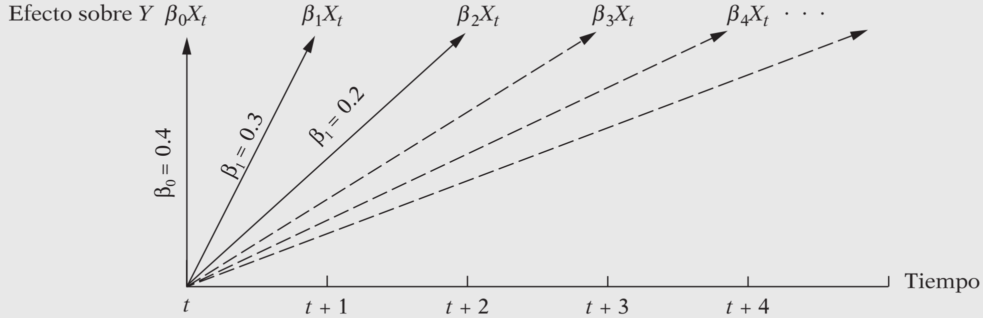



**FIGURA 17.2**
Efecto de un cambio unitario en $X$ sobre $Y$ en el tiempo $t$ y sobre los periodos subsiguientes.



En forma más general, escribiríamos

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \qquad (17.1.2)$$

que es el modelo de rezagos distribuidos con un rezago finito de $k$ periodos. El coeficiente $\beta_0$ se conoce como **multiplicador de corto plazo** o **de impacto** porque da el cambio en el valor medio de $Y$ que sigue a un cambio unitario en $X$ en el mismo periodo.[1] Si el cambio en $X$ se mantiene igual desde el principio, entonces $(\beta_0 + \beta_1)$ da el cambio en (el valor medio de) $Y$ en el periodo siguiente $(\beta_0 + \beta_1 + \beta_2)$ en el que le sigue, y así sucesivamente. Estas sumas parciales se denominan **multiplicadores ínterin**, o **intermedios**. Por último, después de $k$ periodos obtenemos

$$\sum_{i=0}^{k} \beta_i = \beta_0 + \beta_1 + \beta_2 + \cdots + \beta_k = \beta \qquad (17.1.3)$$

que se conoce como **multiplicador de rezagos distribuidos de largo plazo** o **total**, siempre que exista la suma $\beta$ (explicaremos esto más adelante).

Si definimos

$$\beta_i^* = \frac{\beta_i}{\sum \beta_i} = \frac{\beta_i}{\beta} \qquad (17.1.4)$$

obtenemos $\beta_i$ "estandarizado". Las sumas parciales del $\beta_i$ estandarizado dan la proporción del impacto de largo plazo, o total, sentido durante cierto periodo.

De vuelta a la regresión de consumo (17.1.1), vemos que el multiplicador de corto plazo, que no es otra cosa que la propensión marginal a consumir de corto plazo (PMC), es 0.4, mientras que el multiplicador de largo plazo, que es la propensión marginal a consumir de largo plazo, es $0.4 + 0.3 + 0.2 = 0.9$. Es decir, después de un incremento de $1 en el ingreso, el consumidor aumentará su nivel de consumo alrededor de 40 centavos de dólar en el año del aumento, otros 30 centavos en el año siguiente y otros 20 centavos más en el siguiente. El impacto de largo plazo de un incremento de $1 en el ingreso es, entonces, de 90 centavos. Si dividimos cada $\beta_i$ entre 0.9, obtenemos 0.44, 0.33 y 0.23, respectivamente, lo cual indica que 44% del impacto total de un cambio unitario en $X$ sobre $Y$ se siente de inmediato, 77% se siente después de un año, y 100%, al finalizar el segundo año.

**EJEMPLO 17.2**
*Creación de dinero bancario (depósitos a la vista)*

Suponga que el Sistema de la Reserva Federal emite $1 000 de dinero nuevo, el cual entrega al sistema bancario mediante la compra de títulos del gobierno. ¿Cuál será la cantidad total del dinero bancario, o depósitos a la vista, que se generarán en último término?

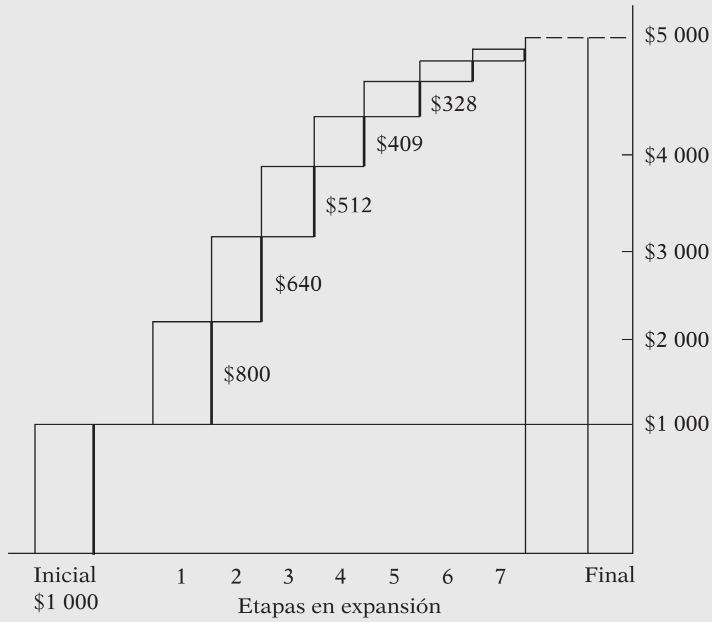
En el contexto del sistema de reservas fraccionales, si suponemos que la ley exige a los bancos retener 20% de las reservas para respaldar los depósitos que ellos crean, entonces, mediante el conocido proceso multiplicador, el total de los depósitos a la vista que se generarán será $\$1\,000[1/(1 - 0.8)] = \$5\,000$. Por supuesto, $5 000 en depósitos a la vista no se crean de la noche a la mañana. El proceso toma tiempo, como se ve esquemáticamente en la figura 17.3.

(*continúa*)

[1] Técnicamente, $\beta_0$ es la derivada parcial de $Y$ respecto de $X_t$, $\beta_1$ respecto de $X_{t-1}$, $\beta_2$ respecto de $X_{t-2}$, y así sucesivamente. Simbólicamente, $\partial Y_t/\partial X_{t-k} = \beta_k$.

**EJEMPLO 17.2**
(*continuación*)

**FIGURA 17.3** Expansión acumulativa en depósitos bancarios (reserva inicial de \$1 000 y requerimientos de encaje de 20%).




**EJEMPLO 17.3**
*Vínculo entre dinero y precios*

De acuerdo con los monetaristas, la inflación es, en esencia, un fenómeno monetario en el sentido de que un incremento continuo en el nivel general de precios se debe a la tasa de expansión en la oferta monetaria que excede en mayor medida la cantidad de dinero realmente demandada por las unidades económicas. Por supuesto, este vínculo entre inflación y cambios en la oferta monetaria no es instantáneo. Algunos estudios demuestran que el rezago entre las dos está en alguna parte entre 3 y alrededor de 20 trimestres. Los resultados de uno de estos estudios se presentan en la tabla 17.1,[2] donde se ve que el efecto de un cambio de 1% en la oferta monetaria M1B (= dinero circulante + depósitos a la vista en las instituciones financieras) se siente durante un periodo de 20 trimestres. El impacto de largo plazo de un cambio de 1% en la oferta monetaria sobre la inflación es de alrededor de 1 ($= \sum m_i$), estadísticamente significativo, mientras que el impacto de corto plazo es de alrededor de 0.04, no significativo, aunque los multiplicadores intermedios por lo general parecen significativos. A propósito, observe que, como $P$ y $M$ se expresan en forma porcentual, las $m_i$ ($\beta_i$ en nuestra notación) dan la elasticidad de $P$ respecto de $M$, es decir, la respuesta porcentual de los precios a un incremento de 1% en la oferta monetaria. Así, $m_0 = 0.041$ significa que para un incremento de 1% en la oferta monetaria, la elasticidad de corto plazo de los precios es de alrededor de 0.04%. La elasticidad de largo plazo es 1.03%, lo cual implica que, en el largo plazo, un incremento de 1% en la oferta monetaria se refleja en apenas alrededor del mismo incremento porcentual en los precios. En resumen, un incremento de 1% en la oferta monetaria se acompaña en el largo plazo por un incremento de 1% en la tasa de inflación.

[2] Keith M. Carlson, "The Lag from Money to Prices", *Review*, Federal Reserve Bank of St. Louis, octubre de 1980, tabla 1, p. 4.

**EJEMPLO 17.3** (*continuación*)

**TABLA 17.1 Estimación de la ecuación dinero-precios: especificación original**

Periodo muestral: 1955–I a 1969–IV: $m_{21} = 0$

$$\dot{P} = \underset{(0.395)}{-0.146} + \sum_{i=0}^{20} m_i \dot{M}_{-i}$$

| | Coef. | $\lvert t \rvert$ | | Coef. | $\lvert t \rvert$ | | Coef. | $\lvert t \rvert$ |
|---|---|---|---|---|---|---|---|---|
| $m_0$ | 0.041 | 1.276 | $m_8$ | 0.048 | 3.249 | $m_{16}$ | 0.069 | 3.943 |
| $m_1$ | 0.034 | 1.538 | $m_9$ | 0.054 | 3.783 | $m_{17}$ | 0.062 | 3.712 |
| $m_2$ | 0.030 | 1.903 | $m_{10}$ | 0.059 | 4.305 | $m_{18}$ | 0.053 | 3.511 |
| $m_3$ | 0.029 | 2.171 | $m_{11}$ | 0.065 | 4.673 | $m_{19}$ | 0.039 | 3.338 |
| $m_4$ | 0.030 | 2.235 | $m_{12}$ | 0.069 | 4.795 | $m_{20}$ | 0.022 | 3.191 |
| $m_5$ | 0.033 | 2.294 | $m_{13}$ | 0.072 | 4.694 | $\sum m_i$ | 1.031 | 7.870 |
| $m_6$ | 0.037 | 2.475 | $m_{14}$ | 0.073 | 4.468 | **Rezago medio** | 10.959 | 5.634 |
| $m_7$ | 0.042 | 2.798 | $m_{15}$ | 0.072 | 4.202 | | | |

$\bar{R}^2$ 0.525 ee 1.066 D.W. 2.00

*Notación:* $\dot{P}$ = tasa de cambio anual compuesta del deflactor del PNB.
$\dot{M}$ = tasa de cambio anual compuesta de M1B.

Fuente: Keith M. Carlson, "The Lag from Money to Prices", *Review*, Federal Reserve Bank of St. Louis, octubre de 1980, tabla 1, p. 4.

**EJEMPLO 17.4**
*Rezagos entre el gasto en IyD y productividad*

La decisión de invertir en gastos de investigación y desarrollo (IyD) y su compensación definitiva en términos de mayor productividad implica un considerable rezago, en realidad formado por diversos rezagos, como "el rezago entre la inversión de los fondos y el momento en el cual los inventos realmente empiezan a aparecer, el rezago entre la invención de una idea o mecanismo y su desarrollo hasta llegar a la etapa en que sea comercialmente aplicable y el rezago que se introduce en razón del proceso de difusión: se requiere tiempo para reemplazar las máquinas viejas por nuevas y mejores".[3]

**EJEMPLO 17.5**
*La curva J de la economía internacional*

Los estudiantes de economía internacional conocen ya la *curva J*, la cual muestra la relación entre el balance comercial y la depreciación de la moneda. Al dar seguimiento a la depreciación de la moneda de un país (por ejemplo, a causa de una devaluación), al principio la balanza comercial se deteriora, pero a la larga mejora, si lo demás se mantiene igual. La curva se muestra en la figura 17.4.

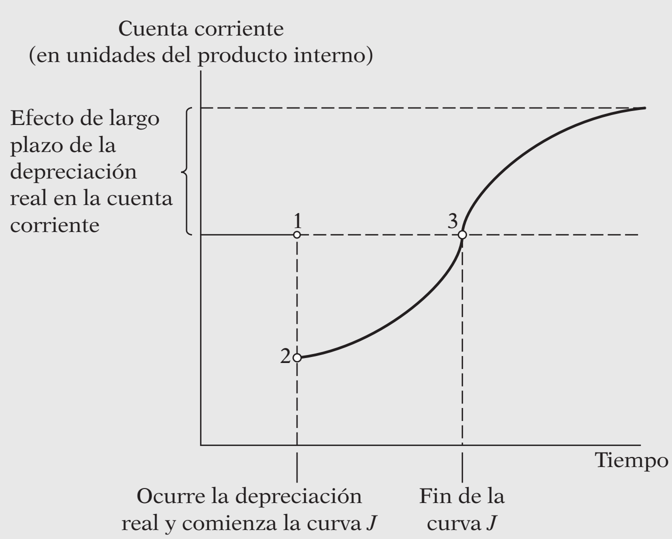



**FIGURA 17.4**
La curva *J*.

Fuente: Paul R. Krugman y Maurice Obstfeld, *International Economics: Theory and Practice*, 3a. ed., Harper Collins, Nueva York, 1994, p. 465.

[3] Zvi Griliches, "Distributed Lags: A Survey", *Econometrica*, vol. 36, núm. 1, enero de 1967, pp. 16-49.

**EJEMPLO 17.6**
*El modelo acelerador de la inversión*

En su forma más sencilla, el principio de aceleración de la teoría de la inversión establece que la inversión es proporcional a los cambios en la producción. De manera simbólica,

$$I_t = \beta(X_t - X_{t-1}) \qquad \beta > 0 \tag{17.1.5}$$

donde $I_t$ es la inversión en el tiempo $t$, $X_t$ es la producción en el tiempo $t$ y $X_{t-1}$ es la producción en el tiempo $(t - 1)$.

Los ejemplos anteriores son una muestra de los rezagos en economía. Sin duda, el lector puede dar diversos ejemplos con base en su propia experiencia.

## 17.2 Razones de los rezagos[4]

Aunque los ejemplos de la sección 17.1 señalan la naturaleza de los fenómenos rezagados, no explican por completo la razón por la cual ocurren. Hay tres razones principales:

1. **Razones psicológicas.** Como resultado de la fuerza del hábito (inercia), la gente no cambia sus hábitos de consumo de inmediato tras una reducción de precios o de un incremento en el ingreso, quizá debido a que el proceso de cambio conlleve alguna desventaja inmediata. Así, quienes de pronto se convierten en millonarios al ganar la lotería quizá no cambien el estilo de vida al cual estaban acostumbrados durante largo tiempo por no saber cómo reaccionar a una ganancia repentina como ésa. Por supuesto, después de un tiempo razonable, aprenden a vivir con su recién adquirida fortuna. Asimismo, la gente puede no saber si un cambio es "permanente" o "transitorio". Así, su reacción a un incremento en su ingreso depende de que el incremento sea permanente o no. Si sólo es un incremento que no se repite y en los periodos siguientes su ingreso retorna al nivel anterior, quizá ahorre la totalidad del incremento, mientras que otra persona en su posición puede decidir "disfrutarlo".
2. **Razones tecnológicas.** Suponga que se reduce el precio del capital relativo al trabajo, de modo que es económicamente factible sustituir mano de obra por capital. Desde luego, la adición de capital toma tiempo (periodo de gestación). Además, si se espera que la caída de precios sea temporal, las empresas pueden no apurarse a sustituir mano de obra por capital, en especial si esperan que luego de la caída temporal el precio del capital tal vez aumente más allá de su nivel anterior. Algunas veces, el conocimiento imperfecto también explica los rezagos. En este momento, el mercado de computadoras personales está lleno de toda clase de computadoras con diversas características y precios. Además, desde su introducción, a finales de la década de los años setenta, los precios de la mayoría de las computadoras personales se han reducido en forma drástica. Como resultado, los posibles consumidores de computadoras personales pueden dudar en comprar hasta que hayan tenido tiempo de revisar las características y los precios de todas las marcas. Además, pueden dudar en comprar ante la expectativa de mayores descensos de precio o de más innovaciones.
3. **Razones institucionales.** Estas razones también contribuyen a los rezagos. Por ejemplo, las obligaciones contractuales pueden impedir que las empresas cambien de una fuente de trabajo o de materias primas a otra. Por ejemplo, quienes colocaron fondos en cuentas de ahorro de largo plazo con término fijo, como uno, tres o siete años, están "atrapados", aunque las condiciones del mercado de dinero ahora permitan rendimientos más altos en otras partes. En forma similar, los empleadores con frecuencia permiten a sus empleados escoger entre diversos planes de seguro de salud, pero sólo se hace una selección, y un empleado no puede cambiarse a otro plan durante al menos un año. Aunque esto puede representar una conveniencia administrativa, el empleado queda comprometido durante un año.

[4] Esta sección se basa en buena medida en Marc Nerlove, "Distributed Lags and Demand Analysis for Agricultural and Other Commodities", *Agricultural Handbook*, núm. 141, Departamento de Agricultura de Estados Unidos, junio de 1958.

Por estas razones, el rezago desempeña un papel central en economía. Esto se refleja en la metodología económica del corto y largo plazos. Por esta razón se dice que las elasticidades precio-ingreso de corto plazo suelen ser menores (en valores absolutos) que las elasticidades correspondientes de largo plazo, o que la propensión marginal a consumir de corto plazo es por lo general menor que la propensión marginal a consumir de largo plazo.

## 17.3 Estimación de modelos de rezagos distribuidos

Ya establecimos que los modelos de rezagos distribuidos desempeñan un papel muy útil en economía, pero ¿cómo se estiman dichos modelos? Suponga que tenemos el siguiente modelo de rezagos distribuidos en una variable explicativa:[5]

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + u_t \qquad \textbf{(17.3.1)}$$

donde no hemos definido la longitud del rezago, es decir, cuán atrás en el pasado deseamos ir. Tal modelo se denomina **modelo de rezagos infinito**, mientras que un modelo del tipo (17.1.2) se denomina **modelo de rezagos distribuidos (rezagos) finito** porque la longitud del rezago $k$ está especificada. Continuaremos con (17.3.1) por su facilidad de manejo matemático, como veremos.[6]

¿Cómo estimamos $\alpha$ y las $\beta$ de (17.3.1)? Podemos adoptar dos enfoques: 1) estimación *ad hoc* y 2) restricciones *a priori* sobre las $\beta$, si suponemos que (las $\beta$) siguen un patrón sistemático. Consideraremos la estimación *ad hoc* en esta sección, y el otro enfoque, en la sección 17.4.

### Estimación *ad hoc* de los modelos de rezagos distribuidos

Como se supone que la variable explicativa $X_t$ es no estocástica (o por lo menos no correlacionada con el término de perturbación $u_t$), igualmente son no estocásticas $X_{t-1}$, $X_{t-2}$, y así sucesivamente. Por consiguiente, en principio, es aplicable el método de mínimos cuadrados ordinarios (MCO) a (17.3.1). Éste es el enfoque de Alt[7] y Tinbergen,[8] quienes sugieren que para estimar (17.3.1) se proceda secuencialmente, es decir, primero la regresión $Y_t$ sobre $X_t$, luego la de $Y_t$ sobre $X_t$ y $X_{t-1}$, después la regresión de $Y_t$ sobre $X_t$, $X_{t-1}$ y $X_{t-2}$, y así sucesivamente. Este procedimiento secuencial se detiene cuando los coeficientes de regresión de las variables rezagadas empiezan a ser estadísticamente insignificantes y/o el coeficiente de por lo menos una variable cambia su signo de positivo a negativo, o viceversa. Según este precepto, Alt efectuó la regresión de $Y$, consumo de gasolina, sobre nuevos pedidos $X$. Con base en información trimestral de 1930 a 1939, los resultados fueron los siguientes:

$$\hat{Y}_t = 8.37 + 0.171X_t$$

$$\hat{Y}_t = 8.27 + 0.111X_t + 0.064X_{t-1}$$

$$\hat{Y}_t = 8.27 + 0.109X_t + 0.071X_{t-1} - 0.055X_{t-2}$$

$$\hat{Y}_t = 8.32 + 0.108X_t + 0.063X_{t-1} + 0.022X_{t-2} - 0.020X_{t-3}$$

[5] Si hay más de una variable explicativa en el modelo, cada variable puede tener un efecto rezagado sobre *Y*. Por simplicidad, sólo suponemos una variable explicativa.

[6] En la práctica, sin embargo, se espera que los coeficientes de los valores distantes de *X* tengan un efecto insignificante sobre *Y*.

[7] F.F. Alt, "Distributed Lags", *Econometrica*, vol. 10, 1942, pp. 113-128.

[8] J. Tinbergen, "Long-Term Foreign Trade Elasticities", *Metroeconomica*, vol. 1, 1949, pp. 174-185.

Alt escogió la segunda regresión como la "mejor" porque en las últimas dos ecuaciones el signo de $X_{t-2}$ no fue estable y en la última ecuación el signo de $X_{t-3}$ fue negativo, lo cual es difícil interpretar en términos económicos.

Aunque la estimación *ad hoc* parece sencilla y discreta, plantea muchas desventajas, como las siguientes:

1. No hay guía *a priori* sobre la longitud máxima que debe tener el rezago.[9]
2. A medida que se estiman rezagos sucesivos, quedan menos grados de libertad, con lo cual se debilita un poco la inferencia estadística. Por lo general, los economistas no tienen la suerte de contar con series largas que les permitan estimar numerosos rezagos.
3. Aún más importante, en la información de series de tiempo económicas, los valores (de rezagos) sucesivos tienden a estar altamente correlacionados; por tanto, sale a relucir la multicolinealidad. Como mencionamos en el capítulo 10, la multicolinealidad genera una estimación imprecisa; es decir, los errores estándar tienden a ser grandes en relación con los coeficientes estimados. Como resultado, con base en el cálculo rutinario de las razones $t$, podemos tender a declarar (erróneamente) que uno o varios coeficientes de los rezagos son estadísticamente no significativos.
4. La búsqueda secuencial de la longitud de los rezagos hace que el investigador pueda incurrir en **la minería de datos**. También, como vimos en la sección 13.4, el nivel de significancia nominal y verdadero para probar hipótesis estadísticas se convierte en un asunto importante en tales búsquedas secuenciales [véase la ecuación (13.4.2)].

En vista de estos problemas, no es muy recomendable el procedimiento de estimación *ad hoc*. Sin duda, deben tenerse en cuenta algunas consideraciones previas o teóricas para tratar las diversas $\beta$ si deseamos resolver el problema de estimación.

## 17.4 Método de Koyck para los modelos de rezagos distribuidos

Koyck propuso un método ingenioso de estimación de los modelos de rezagos distribuidos. Suponga que empezamos con un modelo de rezagos distribuidos infinito (17.3.1). *Si todas las $\beta$ tienen el mismo signo*, Koyck da por hecho que se reducen geométricamente de la siguiente manera.[10]

$$\beta_k = \beta_0 \lambda^k \qquad k = 0, 1, \ldots \tag{17.4.1}$$ [11]

donde $\lambda$, tal que $0 < \lambda < 1$, se conoce como tasa de descenso, o de caída, del rezago distribuido y donde $1 - \lambda$ se conoce como velocidad de ajuste.

Lo que se postula en (17.4.1) es que cada coeficiente $\beta$ sucesivo es numéricamente inferior a cada $\beta$ anterior (esta afirmación se debe a que $\lambda < 1$), lo cual implica que, a medida que se retorna al pasado distante, el efecto de ese rezago sobre $Y_t$ se reduce progresivamente, supuesto muy razonable. Después de todo, se espera que los ingresos actuales y del pasado reciente afecten al gasto de consumo actual con mayor peso que el ingreso en el pasado distante. En la figura 17.5 se ilustra geométricamente el esquema de Koyck.

Como muestra esta figura, el valor del coeficiente del rezago $\beta_k$ depende, aparte del $\beta_0$ común, del valor de $\lambda$. Entre más cerca de 1 esté $\lambda$, más lenta será la tasa de descenso en $\beta_k$, mientras que,

---

[9] Si la longitud del rezago, $k$, está especificada incorrectamente, tendremos que enfrentar el problema de errores de especificación analizado en el capítulo 13. Tenga en mente también la advertencia sobre la **minería de datos**.

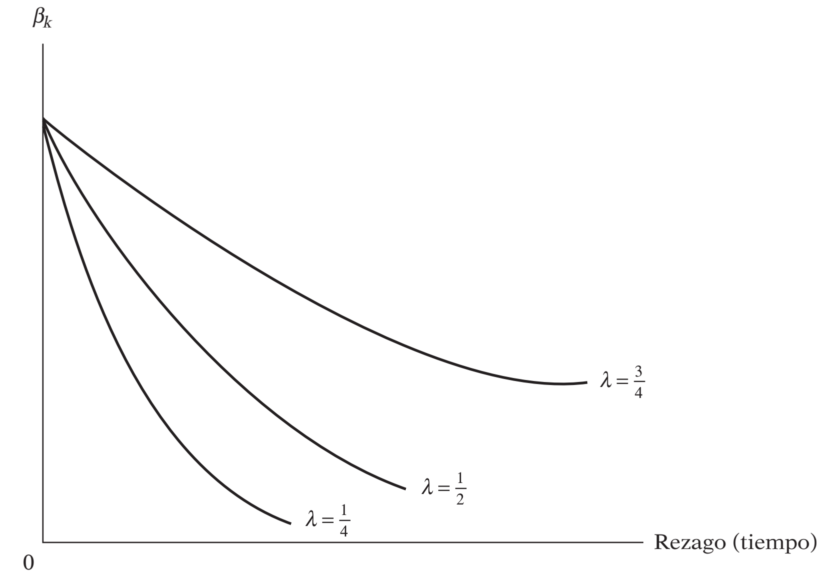
[10] L.M. Koyck, *Distributed Lags and Investment Analysis*, North Holland, Ámsterdam, 1954.

[11] Algunas veces esto se escribe también como

$$\beta_k = \beta_0(1-\lambda)\lambda^k \qquad k = 0, 1, \ldots$$

por las razones dadas en la nota 12.

**FIGURA 17.5**
Esquema de Koyck (distribución geométrica descendente).




entre más cerca esté de cero, más rápido será el descenso en $\beta_k$. En el primer caso, los valores del pasado distante de $X$ ejercerán un impacto considerable sobre $Y_t$, mientras que en el último caso, su influencia sobre $Y_t$ desaparecerá con rapidez. Este patrón se ve claramente en la siguiente ilustración:

| $\lambda$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\cdots$ | $\beta_{10}$ |
|---|---|---|---|---|---|---|---|---|
| 0.75 | $\beta_0$ | $0.75\beta_0$ | $0.56\beta_0$ | $0.42\beta_0$ | $0.32\beta_0$ | $0.24\beta_0$ | $\cdots$ | $0.06\beta_0$ |
| 0.25 | $\beta_0$ | $0.25\beta_0$ | $0.06\beta_0$ | $0.02\beta_0$ | $0.004\beta_0$ | $0.001\beta_0$ | $\cdots$ | 0.0 |

Observe estas características del esquema de Koyck: 1) Al suponer valores no negativos para $\lambda$, Koyck elimina la posibilidad de que las $\beta$ cambien de signo; 2) al suponer que $\lambda < 1$, le da un menor peso a las $\beta$ en el pasado distante que a las actuales; y 3) asegura que la suma de las $\beta$, que proporciona el multiplicador de largo plazo, sea finita, a saber,

$$\sum_{k=0}^{\infty} \beta_k = \beta_0 \left( \frac{1}{1-\lambda} \right) \qquad \textbf{(17.4.2)}^{12}$$

Como resultado de (17.4.1), el modelo de rezagos infinitos (17.3.1) se escribe como

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \cdots + u_t \qquad \textbf{(17.4.3)}$$

Como está planteado, el modelo aún no es adecuado para su fácil estimación, pues un gran número (literalmente infinito) de parámetros quedan aún por estimar y el parámetro $\lambda$ ingresa de

[12] Esto se debe a que

$$\sum \beta_k = \beta_0 (1 + \lambda + \lambda^2 + \lambda^3 + \cdots) = \beta_0 \left( \frac{1}{1-\lambda} \right)$$

porque la expresión en paréntesis en el lado derecho es una serie geométrica infinita cuya suma es $1/(1-\lambda)$ siempre y cuando $0 < \lambda < 1$. A propósito, observe que si $\beta_k$ se define como en la nota 11, $\sum \beta_k = \beta_0 (1-\lambda)/(1-\lambda) = \beta_0$, con lo que se asegura que las ponderaciones $(1-\lambda)\lambda^k$ sumen uno.

forma por completo no lineal: en sentido estricto, el método de análisis de regresión lineal (en parámetros) no puede aplicarse a un modelo de este tipo. Pero ahora Koyck sugiere una forma ingeniosa para lograrlo. Rezaga (17.4.3) en un periodo para obtener

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \beta_0 \lambda X_{t-2} + \beta_0 \lambda^2 X_{t-3} + \cdots + u_{t-1} \tag{17.4.4}$$

Luego multiplica (17.4.4) por λ para obtener

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \beta_0 \lambda^3 X_{t-3} + \cdots + \lambda u_{t-1} \tag{17.4.5}$$

Al restar (17.4.5) de (17.4.3), obtiene

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta_0 X_t + (u_t - \lambda u_{t-1}) \tag{17.4.6}$$

o, al reordenar,

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t \tag{17.4.7}$$

donde $v_t = (u_t - \lambda u_{t-1})$ es un **promedio móvil de $u_t$ y $u_{t-1}$**.

El procedimiento recién descrito se conoce como **transformación de Koyck**. Al comparar (17.4.7) con (17.3.1), vemos la enorme simplificación que logró Koyck. Mientras que antes era preciso estimar $\alpha$ y un número infinito de $\beta$, ahora sólo hay que estimar tres incógnitas: $\alpha$, $\beta_0$ y λ. Ahora bien, no hay razón para esperar multicolinealidad. En cierto sentido, la multicolinealidad se resuelve al reemplazar $X_{t-1}, X_{t-2}, \ldots$, por una variable única, a saber, $Y_{t-1}$. Pero observe las siguientes características de la transformación de Koyck:

1. Empezamos con un modelo de rezagos distribuidos y terminamos con un modelo autorregresivo porque $Y_{t-1}$ aparece como una variable explicativa. Esta transformación muestra la forma como un modelo de rezagos distribuidos se "convierte" en un modelo autorregresivo.
2. Es probable que la aparición de $Y_{t-1}$ cree algunos problemas estadísticos. $Y_{t-1}$, al igual que $Y_t$, es estocástica, lo cual significa que tenemos una variable explicativa estocástica en el modelo. Recuerde que la teoría clásica de mínimos cuadrados se basa en el supuesto de que las variables explicativas son no estocásticas o, en caso de serlo, están distribuidas independientemente del término de perturbación estocástico. Por tanto, debemos averiguar si $Y_{t-1}$ satisface este supuesto. (Volveremos a este punto en la sección 17.8.)
3. En el modelo original (17.3.1), el término de perturbación era $u_t$, mientras que en el modelo transformado es $v_t = (u_t - \lambda u_{t-1})$. Las propiedades estadísticas de $v_t$ dependen de lo que se suponga sobre las propiedades estadísticas de $u_t$, pues, como veremos más adelante, si las $u_t$ originales no están serialmente correlacionadas, las $v_t$ están serialmente correlacionadas. Por consiguiente, tal vez enfrentemos el problema de correlación serial adicional a la variable explicativa estocástica $Y_{t-1}$. Haremos esto en la sección 17.8.
4. La presencia de la $Y$ rezagada viola un supuesto en que se basa la prueba $d$ de Durbin-Watson. Por consiguiente, debemos desarrollar una prueba alterna para verificar la correlación serial en presencia de una $Y$ rezagada. Una alternativa es la **prueba *h* de Durbin**, que analizamos en la sección 17.10.

Como vimos en (17.1.4), las sumas parciales de las $\beta_i$ estandarizadas reflejan la proporción del impacto de largo plazo, o total, sentido durante un cierto periodo. En la práctica, sin embargo, con el **rezago medio** o **mediano** a menudo se caracteriza la naturaleza de la estructura de los rezagos de un modelo de rezagos distribuidos.

## Mediana de los rezagos

La mediana de los rezagos es el tiempo requerido para la primera mitad, o 50%, del cambio total ocurrido en *Y* como consecuencia de un cambio unitario sostenido en *X*. Para el modelo de Koyck, la mediana de los rezagos es la siguiente (véase el ejercicio 17.6):

$$\text{Modelo de Koyck: Mediana de los rezagos} = -\frac{\log 2}{\log \lambda} \tag{17.4.8}$$

Así, si $\lambda = 0.2$, la mediana de rezagos es 0.4306, pero si $\lambda = 0.8$, la mediana de rezagos es 3.1067. En palabras, en el primer caso, 50% del cambio total en *Y* se logra en menos de la mitad de un periodo, mientras que en el último caso, requiere más de 3 periodos para alcanzar el cambio de 50%. Pero este contraste no debe sorprender, pues, como sabemos, entre más alto sea el valor de λ menor será la velocidad del ajuste, y entre menor sea el valor de λ mayor será la velocidad del ajuste.

## Rezago medio

En tanto todas las $\beta_k$ sean positivas, el rezago medio o promedio se define como

$$\text{Rezago medio} = \frac{\sum_0^\infty k\beta_k}{\sum_0^\infty \beta_k} \tag{17.4.9}$$

que es tan sólo el promedio ponderado de todos los rezagos participantes, con los respectivos coeficientes $\beta$ como ponderaciones. En resumen, es un **promedio ponderado rezagado** de tiempo. Para el modelo de Koyck, el rezago medio es (véase el ejercicio 17.7)

$$\text{Modelo de Koyck: Rezago medio} = \frac{\lambda}{1-\lambda} \tag{17.4.10}$$

Así, si $\lambda = \frac{1}{2}$, el rezago medio es 1.

De la exposición anterior, es claro que la mediana y la media de los rezagos sirven como medida resumen de la velocidad con la cual *Y* responde a *X*. En el ejemplo de la tabla 17.1, el rezago medio es alrededor de 11 trimestres, lo que indica que se requiere algún tiempo, en promedio, para que el efecto de los cambios en la oferta monetaria se sienta en los cambios de precios.

**EJEMPLO 17.7**
*Gasto de consumo personal per cápita (GCPC) e ingreso disponible personal per cápita (IDPC)*

Este ejemplo analiza el gasto de consumo personal *per cápita* (GCPC) en relación con el ingreso disponible personal *per cápita* (IDPC) en Estados Unidos de 1959 a 2006; todos los datos están en dólares de 2000. Como ilustración del modelo Koyck, considere los datos de la tabla 17.2. La regresión del GCPC sobre el IDPC y el GCPC rezagado arrojó los resultados que se presentan en la tabla 17.3.

La función de consumo en esta tabla se denomina función de consumo de corto plazo. En breve derivaremos la función de consumo de largo plazo.

Con el valor estimado de λ calculamos los coeficientes de los rezagos distribuidos. Si $\beta 0 \approx 0.2139$, $\beta_1 = (0.2139)(0.7971) \approx 0.1704$, $\beta_2 = (0.2139)(0.7971)^2 \approx 0.0231$, y así sucesivamente, que son multiplicadores de corto y mediano plazos. Por último, con la ecuación (17.4.2) obtenemos el multiplicador de largo plazo, es decir, el efecto total del cambio del ingreso sobre el consumo después de tomar en cuenta todos los efectos rezagados, que en el presente ejemplo es:

$$\sum_0^\infty \beta_k = \beta_0\left(\frac{1}{1-\lambda}\right) = (0.2139)\left(\frac{1}{1-0.7971}\right) \approx 1.0537$$

(*continúa*)

**EJEMPLO 17.7**
(*continuación*)

**TABLA 17.2 GCPC e IDPC, 1959-2006**

| Año | GCPC | IDPC | Año | GCPC | IDPC |
|---|---|---|---|---|---|
| 1959 | 8 776 | 9 685 | 1983 | 15 656 | 17 828 |
| 1960 | 8 873 | 9 735 | 1984 | 16 343 | 19 011 |
| 1961 | 8 873 | 9 901 | 1985 | 17 040 | 19 476 |
| 1962 | 9 170 | 10 227 | 1986 | 17 570 | 19 906 |
| 1963 | 9 412 | 10 455 | 1987 | 17 994 | 20 072 |
| 1964 | 9 839 | 11 061 | 1988 | 18 554 | 20 740 |
| 1965 | 10 331 | 11 594 | 1989 | 18 898 | 21 120 |
| 1966 | 10 793 | 12 065 | 1990 | 19 067 | 21 281 |
| 1967 | 10 994 | 12 457 | 1991 | 18 848 | 21 109 |
| 1968 | 11 510 | 12 892 | 1992 | 19 208 | 21 548 |
| 1969 | 11 820 | 13 163 | 1993 | 19 593 | 21 493 |
| 1970 | 11 955 | 13 563 | 1994 | 20 082 | 21 812 |
| 1971 | 12 256 | 14 001 | 1995 | 20 382 | 22 153 |
| 1972 | 12 868 | 14 512 | 1996 | 20 835 | 22 546 |
| 1973 | 13 371 | 15 345 | 1997 | 21 365 | 23 065 |
| 1974 | 13 148 | 15 094 | 1998 | 22 183 | 24 131 |
| 1975 | 13 320 | 15 291 | 1999 | 23 050 | 24 564 |
| 1976 | 13 919 | 15 738 | 2000 | 23 860 | 25 469 |
| 1977 | 14 364 | 16 128 | 2001 | 24 205 | 25 687 |
| 1978 | 14 837 | 16 704 | 2002 | 24 612 | 26 217 |
| 1979 | 15 030 | 16 931 | 2003 | 25 043 | 26 535 |
| 1980 | 14 816 | 16 940 | 2004 | 25 711 | 27 232 |
| 1981 | 14 879 | 17 217 | 2005 | 26 277 | 27 436 |
| 1982 | 14 944 | 17 418 | 2006 | 26 828 | 28 005 |

*Notas:* GCPC = gasto de consumo personal *per cápita*, en dólares de 2000, ajustados por la inflación.
IDPC = ingreso disponible *per cápita*, en dólares de 2000, ajustados por la inflación.

Fuente: *Economic Report of the President*, 2007, tabla B-31.

**TABLA 17.3**

Variable dependiente: GCPC
Método: Mínimos cuadrados
Muestra (ajustada): 1960-2006
Observaciones incluidas: 47 después de los ajustes

| | Coeficiente | Error estándar | Estadístico $t$ | Prob. |
|---|---|---|---|---|
| $C$ | -252.9190 | 157.3517 | -1.607348 | 0.1151 |
| GCPC | 0.213890 | 0.070617 | 3.028892 | 0.0041 |
| IDPC(-1) | 0.797146 | 0.073308 | 10.87389 | 0.0000 |

| | | | |
|---|---|---|---|
| $R$ cuadrada | 0.998216 | Media de la variable dependiente | 16691.28 |
| $R$ cuadrada ajustada | 0.998134 | Desviación estándar de la variable dependiente | 5205.873 |
| Error estándar de la regresión | 224.8504 | | |
| Suma de cuadrados residual | 2224539. | Criterio de información de Akaike | 13.73045 |
| Log verosimilitud | -319.6656 | Criterio de Schwarz | 13.84854 |
| Estadístico $F$ | 12306.99 | Criterio de Hannan-Quinn | 13.77489 |
| Prob. (estadístico $F$) | 0.000000 | Estadístico Durbin-Watson | 0.961921 |
| | | Durbin $h$ = 3.8269* | |

*Explicamos el cálculo $h$ de Durbin en la sección 17.10.

**EJEMPLO 17.7** (*continuación*)

En palabras, un incremento sostenido de un dólar en el IDPC producirá al final alrededor de 1.05 dólares de aumento en el GCPC, pero el impacto inmediato, o de corto plazo, es de sólo 21 centavos de dólar.

La función de consumo de largo plazo se expresa así:

$$\text{GCPC}_t = -1\,247.1351 + 1.0537\text{IDPC}_t$$

Esto se obtiene al dividir la función de consumo de corto plazo presentada en la tabla 17.3 entre 0.2029 en ambos lados y eliminar el término de IDPC rezagado.[13]

En el largo plazo, la propensión marginal a consumir (PMC) es alrededor de 1. Esto significa que cuando los consumidores tienen tiempo de adaptarse al incremento de un dólar en el IDPC, aumentan el GCPC casi un dólar. En el corto plazo, sin embargo, como muestra la tabla 17.3, la PMC es de sólo 21 centavos. ¿Cuál es la razón de la diferencia entre la PMC de corto y largo plazos?

La respuesta se encuentra en los rezagos mediano y medio. Con $\lambda = 0.7971$, el rezago mediano es:

$$-\frac{\log(2)}{\log \lambda} = -\frac{\log(2)}{\log(0.7971)} = 3.0589$$

y el rezago medio es:

$$\frac{\lambda}{1-\lambda} = 3.9285$$

Al parecer, el GCPC real se ajusta al IDPC con un rezago considerable: recuerde que cuanto mayor sea el valor de $\lambda$ (entre 0 y 1), más tiempo tardará en sentirse el impacto pleno de un cambio en el valor de la variable explicativa en la variable dependiente.

## 17.5 Racionalización del modelo de Koyck: modelo de expectativas adaptativas

Aunque es muy claro, el modelo de Koyck (17.4.7) es *ad hoc,* pues se obtuvo mediante un proceso puramente algebraico; está desprovisto de cualquier soporte teórico. Pero esta falla se compensa si empezamos desde otra perspectiva. Suponga que postulamos el siguiente modelo:

$$Y_t = \beta_0 + \beta_1 X_t^* + u_t \qquad \textbf{(17.5.1)}$$

donde $Y$ = demanda de dinero (balances reales de efectivo)
$X^*$ = tasa de interés normal o esperada de largo plazo o de equilibrio, u óptima
$u$ = término de error

La ecuación (17.5.1) postula que la demanda de dinero es función de la tasa de interés *esperada* (es decir, anticipada).

Como la variable de expectativas $X^*$ no es directamente observable, podemos proponer la siguiente hipótesis sobre la manera de conformar las expectativas:

$$X_t^* - X_{t-1}^* = \gamma(X_t - X_{t-1}^*) \qquad \textbf{(17.5.2)}^{14}$$

[13] En equilibrio, todos los valores de GCPC serán iguales. Por tanto, $\text{GCPC}_t = \text{GCPC}_{t-1}$. Con esta sustitución se obtiene la función de consumo de largo plazo.

[14] Algunas veces el modelo se expresa como

$$X_t^* - X_{t-1}^* = \gamma(X_{t-1} - X_{t-1}^*)$$

donde $\gamma$, tal que $0 < \gamma \leq 1$, se conoce como **coeficiente de expectativas**. La hipótesis (17.5.2) se conoce como hipótesis de **expectativas adaptativas, expectativas progresivas** o **de aprendizaje por error**, popularizada por Cagan[15] y Friedman.[16]

Lo que implica (17.5.2) es que los "agentes económicos adaptarán sus expectativas a la luz de la experiencia pasada y que, en particular, aprenderán de sus errores".[17] De modo más específico, (17.5.2) establece que las expectativas se corrigen cada periodo por una fracción $\gamma$ de la brecha entre el valor actual de la variable y su valor esperado anterior. Así, para el modelo esto significaría que las expectativas sobre tasas de interés son corregidas cada periodo por una fracción $\gamma$ de la discrepancia entre la tasa de interés observada en el periodo actual y lo que fue su valor anticipado en el periodo anterior. Otra forma de plantear esto sería escribir (17.5.2) como

$$X_t^* = \gamma X_t + (1-\gamma)X_{t-1}^* \tag{17.5.3}$$

lo cual muestra que el valor esperado de la tasa de interés en el tiempo $t$ es un promedio ponderado del valor actual de la tasa de interés en el tiempo $t$ y su valor esperado en el periodo anterior, con ponderaciones de $\gamma$ y $1-\gamma$, respectivamente. Si, $\gamma = 1$, $X_t^* = X_t$, lo cual significa que las expectativas se cumplen de inmediato y en forma completa, es decir, en el mismo periodo. Si, por otra parte, $\gamma = 0$, $X_t^* = X_{t-1}^*$, lo que significa que las expectativas son estáticas, es decir, "las condiciones que prevalecen hoy se mantendrán en todos los periodos subsecuentes. Los valores futuros esperados se identifican entonces con los valores actuales".[18]

Al sustituir (17.5.3) en (17.5.1), obtenemos

$$\begin{aligned} Y_t &= \beta_0 + \beta_1[\gamma X_t + (1-\gamma)X_{t-1}^*] + u_t \\ &= \beta_0 + \beta_1\gamma X_t + \beta_1(1-\gamma)X_{t-1}^* + u_t \end{aligned} \tag{17.5.4}$$

Ahora rezague (17.5.1) un periodo, multiplíquelo por $1-\gamma$ y reste el producto de (17.5.4). Tras un reordenamiento algebraico sencillo, obtenemos

$$\begin{aligned} Y_t &= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + u_t - (1-\gamma)u_{t-1} \\ &= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + v_t \end{aligned} \tag{17.5.5}$$

donde $v_t = u_t - (1-\gamma)u_{t-1}$.

Antes de continuar, es preciso advertir sobre la diferencia entre (17.5.1) y (17.5.5). En la primera, $\beta_1$ mide la respuesta promedio de $Y$ ante un cambio unitario en $X^*$, el valor de equilibrio o de largo plazo de $X$. En (17.5.5), por otra parte, $\gamma\beta_1$ mide la respuesta promedio de $Y$ ante un cambio unitario en el valor actual u observado de $X$. Estas respuestas no serán las mismas a menos, por supuesto, que $\gamma = 1$, es decir, que los valores actuales y de largo plazo de $X$ sean los mismos. En la práctica, primero estimamos (17.5.5). Una vez obtenida una estimación de $\gamma$ a partir del coeficiente de la $Y$ rezagada, podemos calcular $\beta_1$ con facilidad, al dividir el coeficiente de $X_t (= \gamma\beta_1)$ entre $\gamma$.

[15] P. Cagan, "The Monetary Dynamics of Hyperinflations", en M. Friedman (ed.), *Studies in the Quantity Theory of Money*, University of Chicago Press, Chicago, 1956.

[16] Milton Friedman, *A Theory of the Consumption Function*, National Bureau of Economic Research, Princeton University Press, Princeton, Nueva Jersey, 1957.

[17] G.K. Shaw, *Rational Expectations: An Elementary Exposition*, St. Martin's Press, Nueva York, 1984, p. 25.

[18] *Ibid.*, pp. 19-20.

La similitud entre el modelo de expectativas adaptativas (17.5.5) y el modelo de Koyck (17.4.7) debe verse fácilmente aunque las interpretaciones de los coeficientes en los dos modelos son diferentes. Observe que, como el modelo de Koyck, el de expectativas adaptativas es autorregresivo, y su término de error es similar al término de error de Koyck. Volveremos a la estimación del modelo de expectativas adaptativas en la sección 17.8, y daremos algunos ejemplos en la sección 17.12. Ahora que delineamos el modelo de expectativas adaptativas (EA), ¿es realista? Es cierto que es más atractivo que el enfoque de Koyck puramente algebraico, pero, ¿es razonable la hipótesis EA? En favor de la hipótesis EA podemos decir lo siguiente:

> Proporciona un medio relativamente simple de diseñar modelos de expectativas en la teoría económica mientras que, a la vez, postula una forma de comportamiento por parte de los agentes económicos que parece eminentemente razonable. La creencia de que la gente aprende de la experiencia es, desde luego, un punto de partida más razonable que el supuesto implícito de que está totalmente desprovista de memoria, característica de la tesis de expectativas estáticas. Además, la afirmación de que las experiencias más distantes ejercen un efecto menor que las experiencias más recientes coincide con el sentido común y parece bien confirmada por la simple observación.[19]

Hasta la llegada de la **hipótesis de expectativas racionales (ER)**, planteada primero por J. Muth y difundida después por Robert Lucas y Thomas Sargent, la hipótesis EA fue muy popular en la economía empírica. Los proponentes de la hipótesis de ER sostienen que la hipótesis de EA es inadecuada porque la formulación de expectativas se basa tan sólo en los valores pasados de una variable,[20] mientras que la hipótesis de ER supone "que los agentes económicos individuales utilizan información actual disponible y relevante en la formación de sus expectativas y no se apoyan únicamente en la experiencia pasada".[21] En resumen, la hipótesis de ER sostiene que "las expectativas son 'racionales' en el sentido de que incorporan con eficiencia *toda* la información disponible en el momento en que se formulan las expectativas"[22] y no sólo la información pasada.

La crítica dirigida por los proponentes de las ER en contra de la hipótesis de EA es bien recibida, aunque hay muchas críticas a la propia hipótesis de ER.[23] Aquí no es el lugar para dejarse enredar con este material reciente. Se puede o no estar de acuerdo con Stephen McNees en que, "en el mejor de los casos, el supuesto de las expectativas adaptativas puede defenderse sólo como una 'hipótesis de trabajo' o aproximación a un mecanismo de formación de expectativas más complejo, quizá cambiante".[24]

**EJEMPLO 17.8**
*Ejemplo 17.7 reconsiderado*

Como el modelo de expectativas adaptativas se basa en la transformación de Koyck, los resultados de la tabla 17.3 también se interpretan en términos de la ecuación (17.5.5). Por consiguiente, $\hat{\gamma}\,\hat{\beta}_0 = -252.9190$; $\hat{\gamma}\,\hat{\beta}_1 = 0.21389$ y $(1 - \hat{\gamma}) = 0.797146$. Así, el coeficiente de expectativas $\hat{\gamma} \approx 0.2028$ y, según el análisis anterior sobre el modelo de EA, podemos decir que alrededor de 20% de la discrepancia entre el IDPC real y el esperado se elimina en el transcurso de un año.

---

[19] *Ibid.*, p. 27.

[20] Al igual que el modelo de Koyck, puede demostrarse que, según EA, las expectativas de una variable son un promedio ponderado exponencialmente de valores pasados de dicha variable.

[21] G.K. Shaw, *op. cit.*, p. 47. Para detalles adicionales de la hipótesis de ER, véase Steven M. Sheffrin, *Rational Expectations*, Cambridge University Press, Nueva York, 1983.

[22] Stephen K. McNees, "The Phillips Curve: Forward- or Backward-Looking?", *New England Economic Review*, julio-agosto de 1979, p. 50.

[23] Hay una evaluación crítica reciente de la hipótesis de ER en Michael C. Lovell, "Test of the Rational Expectations Hypothesis", *American Economic Review*, marzo de 1966, pp. 110-124.

[24] Stephen K. McNees, *op. cit.*, p. 50.

## 17.6 Otra racionalización del modelo de Koyck: el modelo de ajuste de existencias o de ajuste parcial

El modelo de expectativas adaptativas es una forma de racionalizar el modelo de Koyck. Otra racionalización es la de Marc Nerlove con el **modelo de ajuste de existencias** o de **ajuste parcial (MAP)**.[25] Para ilustrar este modelo, considere el **modelo de acelerador flexible** de la teoría económica, que supone un nivel de reservas o existencias de capital de equilibrio, óptimo, deseado, o de largo plazo requerido para generar una producción dada en ciertas condiciones de tecnología, tasa de interés, etc. Por simplicidad, suponga que este nivel de capital deseado $Y_i^*$ es una función lineal de la producción $X$ de la siguiente manera:

$$Y_t^* = \beta_0 + \beta_1 X_t + u_t \tag{17.6.1}$$

Como el nivel deseado de capital no es directamente observable, Nerlove postula la siguiente hipótesis, conocida como **hipótesis de ajuste parcial** o **ajuste de existencias**:

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1}) \qquad \textbf{(17.6.2)}^{26}$$

donde $\delta$, tal que $0 < \delta \leq 1$, se conoce como **coeficiente de ajuste** y donde $Y_t - Y_{t-1} =$ cambio real, y $(Y_t^* - Y_{t-1})$ cambio deseado.

Como $Y_t - Y_{t-1}$, el cambio en las existencias de capital entre dos periodos, no es otra cosa que la inversión (17.6.2) se escribe también como

$$I_t = \delta(Y_t^* - Y_{t-1}) \tag{17.6.3}$$

donde $I_t =$ inversión en el periodo $t$.

La ecuación (17.6.2) postula que el cambio real en las existencias de capital (la inversión) en cualquier momento del tiempo $t$ es alguna fracción $\delta$ del cambio deseado durante ese periodo. Si $\delta = 1$, significa que las reservas de capital real son iguales a las deseadas; es decir, las reservas reales se ajustan al instante (durante el mismo periodo) a las deseadas. Sin embargo, si $\delta = 0$, significa que nada cambia, pues las existencias reales en el tiempo $t$ son las mismas que las observadas en el periodo anterior. De hecho, se espera que $\delta$ se encuentre dentro de estos extremos, pues es probable que el ajuste a las existencias deseadas de capital sea incompleto debido a rigideces, inercia, obligaciones contractuales, etc., de aquí el nombre de **modelo de ajuste parcial**. Observe que el mecanismo de ajuste (17.6.2) también se escribe como

$$Y_t = \delta Y_t^* + (1-\delta)Y_{t-1} \tag{17.6.4}$$

lo cual muestra que las existencias de capital observadas en el tiempo $t$ son un promedio ponderado de las existencias de capital deseado en ese momento y de las existencias de capital observadas en el periodo anterior, con $\delta$ y $(1-\delta)$ como ponderaciones. Ahora, la sustitución de (17.6.1) en (17.6.4) da

$$\begin{aligned} Y_t &= \delta(\beta_0 + \beta_1 X_t + u_t) + (1-\delta)Y_{t-1} \\ &= \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta u_t \end{aligned} \tag{17.6.5}$$

[25] Marc Nerlove, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*, *op. cit.*

[26] Algunos autores no agregan el término estocástico de error $u_t$ a la relación (17.6.1) pero lo añaden a esta relación, en la creencia de que si la primera es de verdad una relación de equilibrio, el término de error no tiene fundamento, mientras que el mecanismo de ajuste puede ser imperfecto y requerir el término de perturbación. A propósito, observe que (17.6.2) se escribe algunas veces también como

$$Y_t - Y_{t-1} = \delta(Y_{t-1}^* - Y_{t-1})$$

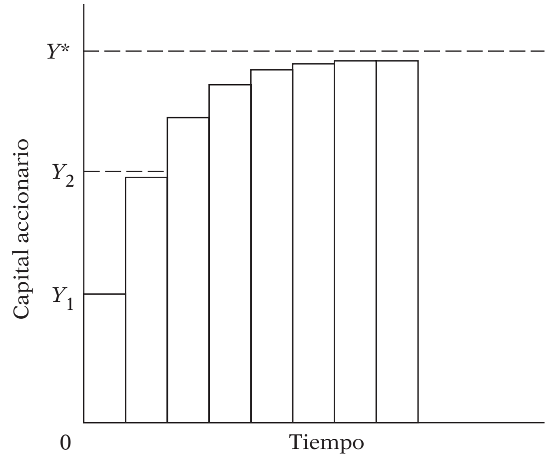



**FIGURA 17.6**
Ajuste gradual de las existencias de capital accionario.

Este modelo se denomina **modelo de ajuste parcial (MAP)**.

Como (17.6.1) representa la demanda de reservas de capital de largo plazo o de equilibrio, (17.6.5) puede denominarse la función de demanda de reservas de capital de corto plazo, pues, en el corto plazo las reservas de capital existentes pueden no ser necesariamente iguales a su nivel en el largo plazo. Una vez estimada la función de corto plazo (17.6.5) y con la estimación del coeficiente de ajuste $\delta$ (del coeficiente de $Y_{t-1}$), se deriva fácilmente la función de largo plazo tan sólo al dividir $\delta\beta_0$ y $\delta\beta_1$ entre $\delta$ y omitir el término rezagado de $Y$, lo cual dará entonces (17.6.1).

En la figura 17.6 presentamos, en términos geométricos el modelo de ajuste parcial.[27] En esta figura, $Y^*$ son las existencias deseadas de capital y $Y_1$ las existencias reales de capital del periodo en curso. Para fines ilustrativos suponga que $\delta = 0.5$. Esto implica que la empresa planea cerrar la mitad de la brecha entre las existencias de capital reales y deseadas cada periodo; así, en el primer periodo, se mueve a $Y_2$, con una inversión igual a $(Y_2 - Y_1)$ que, a su vez, es igual a la mitad de $(Y^* - Y_1)$. En cada periodo subsiguiente, ésta cierra la mitad de la brecha entre las existencias de capital a principios del periodo y las existencias deseadas de capital $Y^*$.

El modelo de ajuste parcial se parece a los modelos de Koyck y de expectativas adaptativas en que es autorregresivo. Sin embargo, tiene un término de perturbación mucho más sencillo: el término de perturbación original $u_t$ multiplicado por una constante $\delta$. Pero tenga en mente que, si bien son similares en apariencia, los modelos de expectativas adaptativas y de ajuste parcial son muy diferentes conceptualmente. El primero se basa en la incertidumbre (sobre el curso futuro de los precios, las tasas de interés, etc.), mientras que el último se debe a rigideces técnicas o institucionales, a la inercia, al costo del cambio, etc. Sin embargo, en teoría ambos modelos son mucho más sólidos que el modelo de Koyck.

Como en apariencia los modelos de expectativas adaptativas y de ajuste parcial son indistinguibles, el coeficiente $\gamma$ de 0.2028 del modelo de expectativas adaptativas también se interpreta como coeficiente $\delta$ del modelo de ajuste de existencias si suponemos que el último modelo es operativo en el presente caso (es decir, es el GCPC deseado o esperado el que está relacionado linealmente con el IDPC actual).

El punto importante es que el modelo de Koyck, el de expectativas adaptativas y el de ajuste de existencias —aparte de la diferencia en la apariencia del término de error— al final producen el mismo modelo estimado; así, los investigadores deben tener gran cuidado al decir al lector qué modelo utilizan y por qué. Por tanto, los investigadores deben especificar el soporte teórico de sus modelos.

[27] Adaptado de la figura 7.4 de Rudiger Dornbusch y Stanley Fischer, *Macroeconomics*, 3a. ed., McGraw-Hill, Nueva York, 1984, p. 216.

## *17.7 Combinación de los modelos de expectativas adaptativas y de ajuste parcial

Considere el siguiente modelo:

$$Y_t^* = \beta_0 + \beta_1 X_t^* + u_t \tag{17.7.1}$$

donde $Y_t^*$ = existencias de capital deseadas y $X_t^*$ = nivel de producción esperado.

Como ni $Y_t^*$ ni $X_t^*$ son directamente observables, podemos utilizar el mecanismo de ajuste parcial para $Y_t^*$ y el modelo de expectativas adaptativas para $X_t^*$ a fin de llegar a la siguiente ecuación de estimación (véase el ejercicio 17.2):

$$\begin{aligned} Y_t &= \beta_0\delta\gamma + \beta_1\delta\gamma X_t + [(1-\gamma)+(1-\delta)]Y_{t-1} \\ &\quad -(1-\delta)(1-\gamma)Y_{t-2} + [\delta u_t - \delta(1-\gamma)u_{t-1}] \\ &= \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + v_t \end{aligned} \tag{17.7.2}$$

donde $v_t = \delta\,[u_t - (1-\gamma)u_{t-1}]$. Este modelo es también autorregresivo y la única diferencia respecto del modelo de expectativas puramente adaptativas consiste en que $Y_{t-2}$ aparece junto con $Y_{t-1}$ como variable explicativa. Al igual que los modelos de Koyck y de EA, el término de error en (17.7.2) sigue un proceso de promedios móviles. Otra característica de este modelo es que, aunque el modelo es lineal en las $\alpha$, no es lineal en los parámetros originales.

Una aplicación conocida de (17.7.1) es la hipótesis del ingreso permanente de Friedman, que plantea que el consumo "permanente" o de largo plazo es una función del ingreso "permanente" o de largo plazo.[28]

La estimación de (17.7.2) presenta los mismos problemas de estimación que el modelo de Koyck o que el modelo de EA, pues todos estos modelos son autorregresivos con estructuras similares de error. Además (17.7.2) implica algunos problemas de estimación no lineal que consideraremos brevemente en el ejercicio 17.10 pero no trataremos a fondo en este libro.

## 17.8 Estimación de modelos autorregresivos

Del análisis realizado hasta el momento tenemos los tres modelos siguientes:

*Koyck*

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + v_t \tag{17.4.7}$$

*Expectativas adaptativas*

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + [u_t - (1-\gamma)u_{t-1}] \tag{17.5.5}$$

*Ajuste parcial*

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta u_t \tag{17.6.5}$$

---

* Opcional.

[28] Milton Friedman, *A Theory of Consumption Function*, Princeton University Press, Princeton, Nueva Jersey, 1957.

Todos estos modelos tienen la forma común siguiente:

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + v_t \tag{17.8.1}$$

es decir, todos son autorregresivos por naturaleza. Por consiguiente, debemos ver ahora el problema de estimación de dichos modelos, porque los mínimos cuadrados clásicos pueden no ser aplicables directamente a ellos. **La razón es doble: la presencia de variables explicativas estocásticas y la posibilidad de correlación serial**.

Ahora, como ya mencionamos, para aplicar la teoría clásica de mínimos cuadrados debemos demostrar que la variable explicativa estocástica $Y_{t-1}$ está distribuida independientemente del término de perturbación $v_t$. Para determinar si esto es así, es esencial conocer las propiedades de $v_t$. Si suponemos que el término de perturbación original $u_t$ satisface todos los supuestos clásicos, como $E(u_t) = 0$, var $(u_t) = \sigma^2$ (supuesto de homoscedasticidad) y cov $(u_t, u_{t+s}) = 0$ para $s \neq 0$ (supuesto de no autocorrelación), $v_t$ puede no heredar todas estas propiedades. Considere, por ejemplo, el término de error en el modelo de Koyck, $v_t = (u_t - \lambda u_{t-1})$. Con los supuestos sobre $u_t$, podemos demostrar fácilmente que $v_t$ está serialmente correlacionada porque

$$E(v_t v_{t-1}) = -\lambda\sigma^2 \tag{17.8.2}$$ [29]

que es diferente de cero (a menos de que λ resulte cero). Y como $Y_{t-1}$ aparece en el modelo de Koyck como variable explicativa, está sujeta a estar correlacionada con $v_t$ (mediante la presencia de $u_{t-1}$ en él). De hecho, podemos demostrar que

$$\text{cov}\,[Y_{t-1}, (u_t - \lambda u_{t-1})] = -\lambda\sigma^2 \tag{17.8.3}$$

que es lo mismo que (17.8.2). El lector puede verificar que sucede lo mismo respecto del modelo de expectativas adaptativas.

¿Cuál es la implicación de encontrar que en el modelo de Koyck, al igual que en el modelo de expectativas adaptativas, la variable explicativa estocástica $Y_{t-1}$ está correlacionada con el término de error $v_t$? Como ya vimos, **si una variable explicativa en un modelo de regresión está correlacionada con el término de perturbación estocástico, los estimadores MCO no sólo están sesgados sino que, además, no son siquiera consistentes; es decir, aunque el tamaño de la muestra aumente indefinidamente, los estimadores no se aproximarán a sus valores poblacionales verdaderos.**[30] **Por consiguiente, la estimación de los modelos de Koyck y de expectativas adaptativas mediante el procedimiento usual de MCO puede producir resultados muy erróneos**.

Sin embargo, el modelo de ajuste parcial es diferente. En este modelo, $v_t = \delta u_t$, donde $0 < \delta \leq 1$. Por consiguiente, si $u_t$ satisface los supuestos del modelo clásico de regresión lineal dados antes, igual lo hará $\delta u_t$. Por tanto, la estimación por MCO del modelo de ajuste parcial dará estimaciones consistentes aunque las estimaciones tiendan a sesgarse (en muestras finitas o pequeñas).[31] Por intuición, la razón de la consistencia es: aunque $Y_{t-1}$ depende de $u_{t-1}$ y de

---

[29] $E(v_t v_{t-1}) = E(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2})$

$= -\lambda E(u_{t-1})^2$ porque se supone que las covarianzas entre las $u$ son cero

$= -\lambda\sigma^2$

[30] La prueba trasciende el alcance de este libro y se encuentra en Griliches, *op. cit.*, pp. 36-38. Sin embargo, véase en el capítulo 18 un esquema de la prueba en otro contexto. Véase también Asatoshi Maeshiro, "Teaching Regression with a Lagged Dependent Variable and Autocorrelated Disturbances", *The Journal of Economic Education*, invierno de 1996, vol. 27, núm. 1, pp. 72-84.

[31] Hay una prueba en J. Johnston, *Econometric Methods*, 3a. ed., McGraw-Hill, Nueva York, 1984, pp. 360-362. Véase además H.E. Doran y J.W.B. Guise, *Single Equation Methods in Econometrics: Applied Regression Analysis*, University of New England Teaching Monograph Series 3, Armidale, Nueva Gales del Sur, Australia, 1984, pp. 236-244.

todos los términos de perturbación anteriores, no está relacionada con el término de error actual $u_t$. Por consiguiente, siempre que $u_t$ sea serialmente independiente, $Y_{t-1}$ también será independiente o por lo menos no estará correlacionada con $u_t$, lo que satisface un supuesto importante de MCO, a saber, la no correlación entre la(s) variable(s) explicativa(s) y el término de perturbación estocástico.

Aunque la estimación por MCO del modelo de ajuste de existencias, o parcial, proporciona una estimación consistente debido a la estructura simple del término de error en un modelo así, no debemos suponer que puede aplicarse en lugar del modelo de Koyck o de expectativas adaptativas.[32] Se aconseja al lector no hacerlo. Un modelo debe seleccionarse con base en consideraciones teóricas sólidas, no porque genere una estimación estadística sencilla. Todo modelo debe considerarse por sus propios méritos, con la debida atención a las perturbaciones estocásticas que aparecen en ellos. Si en modelos como el de Koyck o el de expectativas adaptativas no es posible aplicar directamente MCO, se deben diseñar métodos para resolver el problema de estimación. Existen otros métodos de estimación, aunque algunos son computacionalmente tediosos. En la siguiente sección veremos uno de estos métodos.

## 17.9 Método de variables instrumentales (VI)

La razón por la cual MCO no es aplicable al modelo de Koyck o de expectativas adaptativas es que la variable explicativa $Y_{t-1}$ tiende a estar correlacionada con el término de error $v_t$. Si de alguna manera es posible eliminar esta correlación, se pueden aplicar MCO para obtener estimaciones consistentes, como ya mencionamos. (*Nota:* Habrá algún sesgo de muestra pequeña.) ¿Cómo se consigue esto? Liviatan propuso la siguiente solución.[33]

Supongamos que encontramos una variable para representar $Y_{t-1}$ muy correlacionada con $Y_{t-1}$ pero no con $v_t$, donde $v_t$ es el término de error en el modelo de Koyck o en el de expectativas adaptativas. Tal representación se denomina **variable instrumental (VI)**.[34] Liviatan sugiere $X_{t-1}$ como variable instrumental para $Y_{t-1}$ y además que los parámetros de la regresión (17.8.1) se obtengan al resolver las siguientes ecuaciones normales:

$$\begin{aligned}
\sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1 \sum X_t + \hat{\alpha}_2 \sum Y_{t-1} \\
\sum Y_t X_t &= \hat{\alpha}_0 \sum X_t + \hat{\alpha}_1 \sum X_t^2 + \hat{\alpha}_2 \sum Y_{t-1} X_t \\
\sum Y_t X_{t-1} &= \hat{\alpha}_0 \sum X_{t-1} + \hat{\alpha}_1 \sum X_t X_{t-1} + \hat{\alpha}_2 \sum Y_{t-1} X_{t-1}
\end{aligned} \tag{17.9.1}$$

Observe que si aplicásemos MCO directamente a (17.8.1), las ecuaciones normales de MCO usuales serían (véase la sección 7.4)

$$\begin{aligned}
\sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1 \sum X_t + \hat{\alpha}_2 \sum Y_{t-1} \\
\sum Y_t X_t &= \hat{\alpha}_0 \sum X_t + \hat{\alpha}_1 \sum X_t^2 + \hat{\alpha}_2 \sum Y_{t-1} X_t \\
\sum Y_t Y_{t-1} &= \hat{\alpha}_0 \sum Y_{t-1} + \hat{\alpha}_1 \sum X_t Y_{t-1} + \hat{\alpha}_2 \sum Y_{t-1}^2
\end{aligned} \tag{17.9.2}$$

La diferencia entre los dos conjuntos de ecuaciones normales debe verse fácilmente. Liviatan demostró que las $\alpha$ estimadas a partir de (17.9.1) son consistentes mientras que las estimadas de

[32] Asimismo, como menciona J. Johnston (*op. cit.*, p. 350), "[el] patrón de ajuste [sugerido por el modelo de ajuste parcial]. . . algunas veces puede no ser razonable".

[33] N. Liviatan, "Consistent Estimation of Distributed Lags", *International Econometric Review*, vol. 4, enero de 1963, pp. 44-52.

[34] Estas variables instrumentales son frecuentes en modelos de ecuaciones simultáneas (véase el capítulo 20).

(17.9.2) pueden no serlo porque $Y_{t-1}$ y $v_t\,[= u_t - \lambda u_{t-1}$ o $u_t - (1-\gamma)u_{t-1}]$ pueden estar correlacionadas, mientras que $X_t$ y $X_{t-1}$ no están correlacionadas con $v_t$. (¿Por qué?)

Aunque en la práctica es fácil aplicarla una vez encontrada una variable representante apropiada, es probable que la técnica de Liviatan presente el problema de multicolinealidad debido a que es factible que $X_t$ y $X_{t-1}$, que forman parte de las ecuaciones normales de (17.9.1), estén muy correlacionadas (como mencionamos en el capítulo 12, es habitual que la mayoría de las series de tiempo económicas presenten un alto grado de correlación entre valores sucesivos). La implicación, entonces, es que, si bien el procedimiento de Liviatan genera estimaciones consistentes, es probable que los estimadores sean ineficientes.[35]

Antes de continuar, la pregunta obvia es ¿cómo encontrar una "buena" variable representante para $Y_{t-1}$, de manera que, aunque esté muy correlacionada con $Y_{t-1}$, no lo esté con $v_t$? Hay algunas sugerencias en la bibliografía, que consideraremos aquí en forma de ejercicio (véase el ejercicio 17.5). Pero debe afirmarse que no siempre es fácil encontrar buenas variables representantes, en cuyo caso el método de VI es de poca utilidad práctica y puede ser necesario recurrir a las técnicas de estimación de máxima verosimilitud, que trascienden el alcance de este libro.[36]

¿Existe alguna prueba para averiguar si el (las) variable(s) instrumental(es) es (son) válida(s)? Dennis Sargan elaboró una prueba, la **prueba SARG**, con este propósito, que describimos en el apéndice 17A, sección 17A.1.

## 17.10 Detección de autocorrelación en modelos autorregresivos: prueba *h* de Durbin

Como vimos, la probable correlación serial en los errores $v_t$ complica un poco el problema de estimación en el modelo autorregresivo: en el modelo de ajuste de existencias, el término de error $v_t$ no tenía correlación serial (de primer orden) cuando el término de error $u_t$ en el modelo original no estaba serialmente correlacionado, mientras que en los modelos de Koyck y de expectativas adaptativas, $v_t$ estaba serialmente correlacionado aunque $u_t$ fuera serialmente independiente. La pregunta entonces es ¿cómo saber si hay correlación serial en el término de error que aparece en los modelos autorregresivos?

Como vimos en el capítulo 12, el estadístico $d$ de Durbin-Watson no sirve para detectar correlación serial (de primer orden) en modelos autorregresivos porque el valor $d$ calculado en tales modelos por lo general tiende a 2, que es el valor de $d$ esperado en una secuencia verdaderamente aleatoria. En otras palabras, si calculamos el estadístico $d$ para tales modelos en la forma habitual, se forma un sesgo inherente que impide descubrir la correlación serial (de primer orden). A pesar de esto, muchos investigadores calculan el valor $d$ por falta de algo mejor. Sin embargo, hace poco Durbin mismo propuso una prueba de muestras grandes para la correlación serial de primer orden en modelos autorregresivos.[37] Esta prueba se llama **estadístico *h***.

Ya analizamos la prueba $h$ de Durbin en el ejercicio 12.36. Por conveniencia, reproducimos el estadístico $h$ (con un ligero cambio en la notación):

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\text{var}\,(\hat{\alpha}_2)]}} \qquad \textbf{(17.10.1)}$$

[35] Para ver la forma de mejorar la eficiencia de los estimadores, consulte Lawrence R. Klien, *A Textbook of Econometrics*, 2a. ed., Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1974, p. 99. Véase también William H. Greene, *Econometric Analysis*, Macmillan, 2a. ed., Nueva York, 1993, pp. 535-538.

[36] Hay un análisis condensado de los métodos de MV en J. Johnston, *op. cit.*, pp. 366-371, al igual que en el apéndice 4A y el apéndice 15A.

[37] J. Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, vol. 38, 1970, pp. 410-421.

donde $n$ = tamaño de la muestra, var($\hat{\alpha}_2$) = varianza del coeficiente de la variable rezagada $Y_t\,(=Y_{t-1})$ en (17.8.1), y $\hat{\rho}$ es la estimación de $\rho$, la correlación serial de primer orden analizada en el capítulo 12.

Como mencionamos en el ejercicio 12.36, para una muestra grande, Durbin demostró que, según la hipótesis nula de que $\rho = 0$, el estadístico $h$ de la ecuación (17.10.1) sigue la distribución normal estandarizada. Es decir,

$$h_{\text{asin}} \sim N(0, 1) \tag{17.10.2}$$

donde asin significa asintóticamente.

En la práctica, como vimos en el capítulo 12, $\rho$ se calcula de la siguiente manera:

$$\hat{\rho} \approx 1 - \frac{d}{2} \tag{17.10.3}$$

Es interesante observar que aunque la $d$ de Durbin no sirve para probar la autocorrelación en los modelos autorregresivos, sí puede emplearse como dato para determinar el estadístico $h$.

Ilustraremos el uso del estadístico $h$ con el ejemplo 17.7. En dicho ejemplo, $n = 47$, $\hat{\rho} \approx (1 - d/2) = 0.5190$ (*Nota:* $d = 0.9619$), y var($\hat{\alpha}_2$) = var(GCPC$_{t-1}$) = $(0.0733)^2 = 0.0053$. Al sustituir esos valores en (17.10.1), obtenemos

$$h = 0.5190\sqrt{\frac{47}{1 - 47(0.0053)}} = 4.1061 \tag{17.10.4}$$

Como este valor $h$ sigue una distribución normal estandarizada según la hipótesis nula, es muy pequeña la probabilidad de obtener un valor $h$ tan alto. Recuerde que la probabilidad de que una variable normal estándar rebase el valor de $\pm 3$ es muy reducida. Así, en este ejemplo, la conclusión es que hay autocorrelación (positiva). Por supuesto, es necesario tener presente que $h$ sigue la distribución normal estandarizada asintóticamente. La muestra de 47 observaciones es razonablemente grande.

Observe estas características del estadístico $h$:

1. No importa cuántas variables $X$ o cuántos valores rezagados de $Y$ se incluyan en el modelo de regresión. Para calcular $h$ debemos considerar sólo la varianza del coeficiente del rezago $Y_{t-1}$.
2. La prueba no es aplicable si [$n$ var($\hat{\alpha}_2$)] es superior a 1. (¿Por qué?) En la práctica, sin embargo, no es usual que esto suceda.
3. Como se trata de una prueba de muestras grandes, su aplicación en muestras pequeñas no se justifica del todo, como demuestran Inder[38] y Kiviet.[39] Se ha sugerido que la prueba de Breusch-Godfrey (BG), también conocida como prueba del multiplicador de Lagrange, analizada en el capítulo 12, es estadísticamente más potente, no sólo en las muestras grandes, sino también en muestras finitas, o pequeñas, y, por consiguiente, es preferible a la prueba $h$.[40]

La conclusión basada en la prueba $h$ que el modelo sufre de autocorrelación se confirma con la prueba de Breusch-Godfrey (BG), que se muestra en la ecuación (12.6.17). Con los siete valores rezagados de los residuos estimados de la regresión de la tabla 17.3, la prueba BG ilustrada en

[38] B. Inder, "An Approximation to the Null Distribution of the Durbin-Watson Statistic in Models Containing Lagged Dependent Variables", *Econometric Theory*, vol. 2, núm. 3, 1986, pp. 413-428.

[39] J.F. Kiviet, "On the Vigour of Some Misspecification Tests for Modelling Dynamic Relationships", *Review of Economic Studies*, vol. 53, núm. 173, 1986, pp. 241-262.

[40] Gabor Korosi, Laszlo Matyas e Istvan P. Szekely, *Practical Econometrics*, Ashgate, Brookfield, Vermont, 1992, p. 92.

**TABLA 17.4**

Variable dependiente: GCP
Método: Mínimos cuadrados
Muestra (ajustada): 1960-2006
Observaciones incluidas: 47 después de los ajustes
Errores estándar CHA Newey-West y covarianza (truncamiento de rezago = 3)

| | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| *C* | -252.9190 | 168.4610 | -1.501350 | 0.1404 |
| IDPC | 0.213890 | 0.051245 | 4.173888 | 0.0000 |
| GCPC (-1) | 0.797146 | 0.051825 | 15.38148 | 0.0000 |

| | | | |
|---|---|---|---|
| *R* cuadrada | 0.998216 | Media de la variable dependiente | 16691.28 |
| *R* cuadrada ajustada | 0.998134 | Desviación estándar de la variable dependiente | 5205.873 |
| Error estándar de la regresión | 224.8504 | Criterio de información de Akaike | 13.73045 |
| Suma de cuadrados residual | 2224539. | Criterio de Schwarz | 13.84854 |
| Log verosimilitud | -319.6656 | Criterio de Hannan-Quinn | 13.77489 |
| Estadístico *F* | 12306.99 | Estadístico Durbin-Watson | 0.961921 |
| Prob. (estadístico *F*) | 0.000000 | | |

la ecuación (12.6.18) obtuvo un valor $\chi^2$ de 15.3869. Para siete grados de libertad (el número de residuos rezagados en la prueba BG), la probabilidad de obtener un valor ji cuadrada de 15.38 o mayor es de aproximadamente 3%, que es muy baja.

Por esta razón es necesario corregir los errores estándar que aparecen en la tabla 17.3, lo cual se efectúa con el procedimiento CHA de Newey-West, estudiado en el capítulo 12. Los resultados se presentan en la tabla 17.4.

Al parecer, MCO subestima los errores estándar de los coeficientes de regresión.

## 17.11 Ejemplo numérico: demanda de dinero en Canadá de I-1979 a IV-1988

Para ilustrar la utilización de los modelos que hemos visto hasta este momento, tenga en cuenta una de las primeras aplicaciones empíricas, a saber, la demanda de dinero (o saldos reales de efectivo). En particular, considere el siguiente modelo:[41]

$$M_t^* = \beta_0 R_t^{\beta_1} Y_t^{\beta_2} e^{u_t} \qquad \textbf{(17.11.1)}$$

donde $M_t^*$ = demanda de dinero (saldos reales de efectivo) deseada, o de largo plazo
$R_t$ = tasa de interés a largo plazo, %
$Y_t$ = ingreso nacional real agregado

Para la estimación estadística (17.11.1) se expresa convenientemente en forma logarítmica como

$$\ln M_t^* = \ln \beta_0 + \beta_1 \ln R_t + \beta_2 \ln Y_t + u_t \qquad \textbf{(17.11.2)}$$

[41] Para un modelo similar, véase Gregory C. Chow, "On the Long-Run and Short-Run Demand for Money", *Journal of Political Economy*, vol. 74, núm. 2, 1966, pp. 111-131. Observe que una ventaja de la función multiplicativa es que los exponentes de las variables proporcionan estimaciones directas de las elasticidades (véase el capítulo 6).

Como la variable de demanda deseada no es observable directamente, suponga la hipótesis de ajuste de existencias, a saber,

$$\frac{M_t}{M_{t-1}} = \left(\frac{M_t^*}{M_{t-1}}\right)^{\delta} \qquad 0 < \delta \leq 1 \qquad \textbf{(17.11.3)}$$

La ecuación (17.11.3) establece que un porcentaje constante (¿por qué?) de la discrepancia entre los saldos reales de efectivo observados y los deseados se elimina en un solo periodo (año). En forma logarítmica, la ecuación (17.11.3) se expresa como

$$\ln M_t - \ln M_{t-1} = \delta(\ln M_t^* - \ln M_{t-1}) \qquad \textbf{(17.11.4)}$$

Al sustituir $\ln M_t^*$ de (17.11.2) en la ecuación (17.11.4) y reordenar, obtenemos

$$\ln M_t = \delta \ln \beta_0 + \beta_1 \delta \ln R_t + \beta_2 \delta \ln Y_t + (1 - \delta) \ln M_{t-1} + \delta u_t \qquad \textbf{(17.11.5)}^{42}$$

que puede denominarse la función de demanda de dinero de corto plazo. (¿Por qué?)

Como una ilustración de la demanda de saldos reales de efectivo de corto y largo plazos, considere los datos de la tabla 17.5. Tales datos trimestrales pertenecen a Canadá, de 1979 a 1988. Las variables se definen de la siguiente forma: *M* [como se definió por la oferta de dinero M1, en dólares canadienses (C$), millones], *P* (deflactor de precio implícito, 1981 = 100), PIB a precios constantes de 1981 (C$, millones) y *R* (tasa de interés corporativa preferencial a 90 días, %).[43] *P* tuvo un efecto de deflación sobre M1, a fin de tener las cifras de los saldos de efectivo reales. *A priori*, se espera que la demanda de dinero real tenga una relación positiva con el PIB (efecto de ingreso positivo) y una negativa con *R* (mientras mayor sea la tasa de interés, mayor será el costo de oportunidad de poseer dinero, en vista de que el dinero M1 paga muy poco interés, si llega a pagar alguno).

Los resultados de la regresión fueron los siguientes:[44]

$$\begin{aligned}
\widehat{\ln M_t} &= 0.8561 \quad - \quad 0.0634 \ln R_t - \quad 0.0237 \ln \text{PIB}_t + \quad 0.9607 \ln M_{t-1} \\
\text{ee} &= (0.5101) \qquad (0.0131) \qquad\quad (0.0366) \qquad\quad (0.0414) \\
t &= (1.6782) \quad (-4.8134) \qquad (-0.6466) \qquad (23.1972) \\
& R^2 = 0.9482 \qquad d = 2.4582 \qquad F = 213.7234 \qquad \textbf{(17.11.6)}
\end{aligned}$$

La función de demanda de corto plazo estimada muestra que la elasticidad del interés de corto plazo tiene el signo apropiado y que es estadísticamente muy significativa, debido a que su valor *p* es casi cero. Sorprendentemente, la elasticidad del ingreso de corto plazo resulta negativa, a pesar de que desde un punto de vista estadístico no es diferente de cero. El coeficiente de ajuste es $\delta = (1 - 0.9607) = 0.0393$, lo cual implica que sólo cerca de 4% de la discrepancia entre el saldo de efectivo real y el deseado se elimina en un trimestre, un ajuste muy lento.

[42] A propósito, observe que este modelo es en esencia no lineal en los parámetros. Por consiguiente, aunque MCO puede generar una estimación insesgada de, por ejemplo, $\beta_1\delta$ tomado en conjunto, puede no dar estimaciones insesgadas de $\beta_1$ y $\delta$ individualmente, sobre todo si la muestra es pequeña.

[43] Estos datos provienen de B. Bhaskar Rao (ed.), *Cointegration for the Applied Economist*, St. Martin's Press, Nueva York, 1994, pp. 210-213. Los datos originales abarcan del primer trimestre de 1956 al cuarto de 1988, pero para propósitos de ejemplificación comenzamos el análisis desde el primer trimestre de 1979.

[44] Observe esta característica de los errores estándar estimados. El error estándar de, por ejemplo, el coeficiente del ln $R_t$ se refiere al error estándar de $\widehat{\beta_1\delta}$, que es un estimador de $\beta_1\delta$. No existe una forma sencilla de obtener los errores estándar de $\hat{\beta}_1$ y $\hat{\delta}$ de manera individual, a partir del error estándar de $\widehat{\beta_1\delta}$, sobre todo si la muestra es relativamente pequeña. No obstante, para muestras grandes, los errores estándar individuales de $\hat{\beta}_1$ y $\hat{\delta}$ se obtienen de manera aproximada, pero los cálculos resultan intrincados. Véase Jan Kmenta, *Elements of Econometrics*, Macmillan, Nueva York, 1971, p. 444.

**TABLA 17.5**
**Dinero, tasa de interés, índice de precios y PIB, Canadá**

Fuente: Rao, *op. cit.*, pp. 210-213.

| Observación | M1 | *R* | *P* | PIB |
|---|---|---|---|---|
| 1979-1 | 22 175.00 | 11.13333 | 0.77947 | 334 800 |
| 1979-2 | 22 841.00 | 11.16667 | 0.80861 | 336 708 |
| 1979-3 | 23 461.00 | 11.80000 | 0.82649 | 340 096 |
| 1979-4 | 23 427.00 | 14.18333 | 0.84863 | 341 844 |
| 1980-1 | 23 811.00 | 14.38333 | 0.86693 | 342 776 |
| 1980-2 | 23 612.33 | 12.98333 | 0.88950 | 342 264 |
| 1980-3 | 24 543.00 | 10.71667 | 0.91553 | 340 716 |
| 1980-4 | 25 638.66 | 14.53333 | 0.93743 | 347 780 |
| 1981-1 | 25 316.00 | 17.13333 | 0.96523 | 354 836 |
| 1981-2 | 25 501.33 | 18.56667 | 0.98774 | 359 352 |
| 1981-3 | 25 382.33 | 21.01666 | 1.01314 | 356 152 |
| 1981-4 | 24 753.00 | 16.61665 | 1.03410 | 353 636 |
| 1982-1 | 25 094.33 | 15.35000 | 1.05743 | 349 568 |
| 1982-2 | 25 253.66 | 16.04999 | 1.07748 | 345 284 |
| 1982-3 | 24 936.66 | 14.31667 | 1.09666 | 343 028 |
| 1982-4 | 25 553.00 | 10.88333 | 1.11641 | 340 292 |
| 1983-1 | 26 755.33 | 9.616670 | 1.12303 | 346 072 |
| 1983-2 | 27 412.00 | 9.316670 | 1.13395 | 353 860 |
| 1983-3 | 28 403.33 | 9.333330 | 1.14721 | 359 544 |
| 1983-4 | 28 402.33 | 9.550000 | 1.16059 | 362 304 |
| 1984-1 | 28 715.66 | 10.08333 | 1.17117 | 368 280 |
| 1984-2 | 28 996.33 | 11.45000 | 1.17406 | 376 768 |
| 1984-3 | 28 479.33 | 12.45000 | 1.17795 | 381 016 |
| 1984-4 | 28 669.00 | 10.76667 | 1.18438 | 385 396 |
| 1985-1 | 29 018.66 | 10.51667 | 1.18990 | 390 240 |
| 1985-2 | 29 398.66 | 9.666670 | 1.20625 | 391 580 |
| 1985-3 | 30 203.66 | 9.033330 | 1.21492 | 396 384 |
| 1985-4 | 31 059.33 | 9.016670 | 1.21805 | 405 308 |
| 1986-1 | 30 745.33 | 11.03333 | 1.22408 | 405 680 |
| 1986-2 | 30 477.66 | 8.733330 | 1.22856 | 408 116 |
| 1986-3 | 31 563.66 | 8.466670 | 1.23916 | 409 160 |
| 1986-4 | 32 800.66 | 8.400000 | 1.25368 | 409 616 |
| 1987-1 | 33 958.33 | 7.250000 | 1.27117 | 416 484 |
| 1987-2 | 35 795.66 | 8.300000 | 1.28429 | 422 916 |
| 1987-3 | 35 878.66 | 9.300000 | 1.29599 | 429 980 |
| 1987-4 | 36 336.00 | 8.700000 | 1.31001 | 436 264 |
| 1988-1 | 36 480.33 | 8.616670 | 1.32325 | 440 592 |
| 1988-2 | 37 108.66 | 9.133330 | 1.33219 | 446 680 |
| 1988-3 | 38 423.00 | 10.05000 | 1.35065 | 450 328 |
| 1988-4 | 38 480.66 | 10.83333 | 1.36648 | 453 516 |

*Notas:* M1 = C$, millones.
*P* = deflactor implícito de precios (1981 = 100).
*R* = tasa de interés corporativa preferencial a 90 días, %.
PIB = C$, millones (precios de 1981).

A fin de volver a obtener la función de demanda de largo plazo (17.11.2), sólo se necesita dividir la función de demanda de corto plazo entre $\delta$ (¿por qué?) y eliminar el término ln $M_{t-1}$. Los resultados son:

$$\widehat{\ln M_t^*} = 21.7888 - 1.6132 \ln R_t - 0.6030 \ln \text{PIB} \qquad \textbf{(17.11.7)}^{45}$$

[45] Note que no presentamos los errores estándar de los coeficientes estimados por las razones dadas en la nota 44.

Como se observa, la elasticidad del interés de largo plazo para la demanda de dinero es sustancialmente más grande (en términos absolutos) que la correspondiente a la elasticidad de corto plazo, lo cual también se cumple para la elasticidad de ingreso, aunque en el presente ejemplo su significancia económica y estadística es dudosa.

Observe que la $d$ de Durbin-Watson estimada es de 2.4582, valor muy cercano a 2, lo cual apoya la observación anterior respecto de que en los modelos autorregresivos la $d$ calculada por lo general está próxima a 2. En consecuencia, no debemos confiar en la $d$ calculada para averiguar si existe correlación serial en los datos. El tamaño de la muestra para este caso es de 40 observaciones, lo cual es razonablemente grande para aplicar la prueba $h$. En este ejemplo, el lector puede verificar que el valor $h$ estimado es $-1.5008$, que no resulta significativo en el nivel de 5%, con lo cual quizá se sugiere que no existe autocorrelación de primer orden en el término de error.

## 17.12 Ejemplos ilustrativos

En esta sección presentamos algunos ejemplos de modelos de rezagos distribuidos para mostrar la forma como los investigadores los utilizan en estudios empíricos.

**EJEMPLO 17.9**
*La Reserva Federal y la tasa de interés real*

Para evaluar el efecto del crecimiento de M1 (circulante + depósitos a la vista) sobre la tasa de interés real de los bonos Aaa, G. J. Santoni y Courtenay C. Stone[46] estimaron, con información mensual, el siguiente modelo de rezagos distribuidos para Estados Unidos.

$$r_t = \text{constante} + \sum_{i=0}^{11} a_i \dot{M}_{t-i} + u_i \tag{17.12.1}$$

donde $r_t$ = índice de rendimiento de los bonos Aaa de Moody menos la tasa de cambio promedio anual en el índice de precios al consumidor ajustado estacionalmente durante los 36 meses anteriores, que sirve como medida de la tasa de interés real, y $M_t$ = crecimiento mensual de $M_1$.

De acuerdo con la "doctrina de la neutralidad del dinero", las variables económicas reales —como producción, empleo, crecimiento económico y tasa de interés real— no reciben influencia permanente del crecimiento monetario y, por consiguiente, no se ven afectadas en esencia por la política monetaria. . . Con este argumento, la Reserva Federal no ejerce influencia permanente alguna sobre la tasa real de interés.[47]

Si esta doctrina es válida, entonces se debe esperar que los coeficientes $a_i$ de los rezagos distribuidos, al igual que su suma, sean estadísticamente no diferentes de cero. Para averiguar si sucede así, los autores estimaron (17.12.1) para dos periodos diferentes, febrero de 1951 a septiembre de 1979 y octubre de 1979 a noviembre de 1982, el último para considerar el cambio en la política monetaria del Banco de la Reserva Federal, la cual desde octubre de 1979 presta mayor atención a la tasa de crecimiento de la oferta monetaria que a la tasa de interés, cuando esta última fue la política en el periodo anterior. Los resultados de su regresión se presentan en la tabla 17.6. Estos resultados parecen apoyar la "doctrina de la neutralidad del dinero", pues durante febrero de 1951 a septiembre de 1979 el crecimiento monetario del periodo en curso, al igual que el rezagado, no tuvieron un efecto estadísticamente significativo sobre la medida de la tasa de interés real. Del mismo modo, durante el último periodo, la doctrina de la neutralidad parece mantenerse, pues $\sum a_i$ no es estadísticamente diferente de cero; sólo el coeficiente $a_1$ es significativo, pero tiene el signo equivocado. (¿Por qué?)

[46] "The Fed and the Real Rate of Interest", *Review*, Federal Reserve Bank of St. Louis, diciembre de 1982, pp. 8-18.

[47] *Ibid.*, p. 15.

**EJEMPLO 17.9**
(*continuación*)

**TABLA 17.6** **Influencia del crecimiento mensual de M1 sobre una medida de la tasa de interés real de los bonos Aaa: febrero de 1951 a noviembre de 1982**

$$r = \text{constante} + \sum_{i=0}^{11} a_i \dot{M}1_{t-1}$$

| | Febrero de 1951 a septiembre de 1979 | | Octubre de 1979 a noviembre de 1982 | |
|---|---|---|---|---|
| | **Coeficiente** | **\|t\|*** | **Coeficiente** | **\|t\|*** |
| Constante | 1.4885† | 2.068 | 1.0360 | 0.801 |
| $a_0$ | −0.00088 | 0.388 | 0.00840 | 1.014 |
| $a_1$ | 0.00171 | 0.510 | 0.03960† | 3.419 |
| $a_2$ | 0.00170 | 0.423 | 0.03112 | 2.003 |
| $a_3$ | 0.00233 | 0.542 | 0.02719 | 1.502 |
| $a_4$ | −0.00249 | 0.553 | 0.00901 | 0.423 |
| $a_5$ | −0.00160 | 0.348 | 0.01940 | 0.863 |
| $a_6$ | 0.00292 | 0.631 | 0.02411 | 1.056 |
| $a_7$ | 0.00253 | 0.556 | 0.01446 | 0.666 |
| $a_8$ | 0.00000 | 0.001 | −0.00036 | 0.019 |
| $a_9$ | 0.00074 | 0.181 | −0.00499 | 0.301 |
| $a_{10}$ | 0.00016 | 0.045 | −0.01126 | 0.888 |
| $a_{11}$ | 0.00025 | 0.107 | −0.00178 | 0.211 |
| $\sum a_i$ | 0.00737 | 0.221 | 0.1549 | 0.926 |
| $\bar{R}^2$ | 0.9826 | | 0.8662 | |
| D-W | 2.07† | | 2.04 | |
| RH01 | 1.27† | 24.536 | 1.40† | 9.838 |
| RH02 | −0.28 | 5.410 | −0.48† | 3.373 |
| NOB | 344. | | 38. | |
| SER ( = SCR) | 0.1548 | | 0.3899 | |

*$|t|$ = valor absoluto de $t$.
† Significativamente diferente de cero en el nivel de 0.05.
Fuente: G. J. Santoni y Courtenay C. Stone, "The Fed and the Real Rate of Interest", *Review*, Federal Reserve Bank of St. Louis, diciembre de 1982, p. 16.

**EJEMPLO 17.10**
*El consumo agregado de corto y largo plazos de Sri Lanka, 1967-1993*

Suponga que el consumo $C$ está relacionado linealmente con el ingreso permanente $X^*$:

$$C_t = \beta_1 + \beta_2 X_t^* + u_t \qquad \textbf{(17.12.2)}$$

Como $X_t^*$ no es observable directamente, necesitamos especificar el mecanismo que genera el ingreso permanente. Suponga que adoptamos la hipótesis de expectativas adaptativas especificada en (17.5.2). Con (17.5.2) y al simplificar, obtenemos la siguiente ecuación de estimación (*cf.* 17.5.5):

$$C_t = \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \qquad \textbf{(17.12.3)}$$

donde $\alpha_1 = \gamma\beta_1$
$\alpha_2 = \gamma\beta_2$
$\alpha_3 = (1 - \gamma)$
$v_t = [u_t - (1 - \gamma)u_{t-1}]$

Como sabemos, $\beta_2$ da la respuesta media del consumo, por ejemplo, a un incremento de un dólar en el ingreso permanente, mientras que $\alpha_2$ da la respuesta media del consumo a un incremento de un dólar en el ingreso actual.

(*continúa*)

**EJEMPLO 17.10** (*continuación*)

A partir de los datos anuales de Sri Lanka de 1967 a 1993 en la tabla 17.5 se obtuvieron los siguientes resultados de la regresión:[48]

$$\begin{aligned} \hat{C} &= 1\,038.403 \quad + \ 0.4043X_t + \ 0.5009C_{t-1} \\ \text{ee} &= (2\,501.455) \quad (0.0919) \quad (0.1213) \\ t &= \quad (0.4151) \quad (4.3979) \quad (4.1293) \\ R^2 &= 0.9912 \quad d = 1.4162 \quad F = 1\,298.466 \end{aligned} \tag{17.12.4}$$

donde $C$ = gasto de consumo privado y $X$ = PIB, ambos a precios constantes. También introdujimos la tasa de interés real en el modelo, pero no fue estadísticamente significativa.

Los resultados muestran que la propensión marginal a consumir de corto plazo (PMC) es igual a 0.4043, lo que sugiere que un incremento de una rupia en el ingreso real observado o actual (como lo mide el PIB) aumentaría el consumo medio en casi 0.40 rupias. Pero si se conserva el aumento en el ingreso, entonces a la larga el PMC, sin el ingreso permanente, sería de $\beta_2 = \gamma\beta_2/\gamma = 0.4043/0.4991 = 0.8100$, o de casi 0.81 rupias. En otras palabras, cuando los consumidores tengan tiempo para ajustarse al cambio de 1 rupia en el ingreso, incrementarán su consumo a final de cuentas en 0.81 rupias.

Ahora, suponga que la función de consumo es

$$C_t^* = \beta_1 + \beta_2 X_t + u_t \tag{17.12.5}$$

En esta formulación, el consumo permanente o de largo plazo $C_t$ es una función lineal del ingreso actual u observado. Como $C_t^*$ no es observable directamente, es preciso acudir al modelo de ajuste parcial (17.6.2). Con este modelo y tras un reordenamiento algebraico, obtenemos

$$\begin{aligned} C_t &= \delta\beta_1 + \delta\beta_2 X_t + (1-\delta)C_{t-1} + \delta u_t \\ &= \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \end{aligned} \tag{17.12.6}$$

En apariencia, este modelo no se diferencia del modelo de expectativas adaptativas (17.12.3). Por consiguiente, los resultados de la regresión en (17.12.4) son por igual aplicables aquí. Sin embargo, hay una gran diferencia en la interpretación de los dos modelos, para no mencionar el problema de estimación asociado al modelo autorregresivo y quizá al serialmente correlacio-

**TABLA 17.7**
**Gasto de consumo privado y PIB, Sri Lanka**

Fuente: Véase la nota 48.

| Observación | GACP | PIB | Observación | GACP | PIB |
|---|---|---|---|---|---|
| 1967 | 61 284 | 78 221 | 1981 | 120 477 | 152 846 |
| 1968 | 68 814 | 83 326 | 1982 | 133 868 | 164 318 |
| 1969 | 76 766 | 90 490 | 1983 | 148 004 | 172 414 |
| 1970 | 73 576 | 92 692 | 1984 | 149 735 | 178 433 |
| 1971 | 73 256 | 94 814 | 1985 | 155 200 | 185 753 |
| 1972 | 67 502 | 92 590 | 1986 | 154 165 | 192 059 |
| 1973 | 78 832 | 101 419 | 1987 | 155 445 | 191 288 |
| 1974 | 80 240 | 105 267 | 1988 | 157 199 | 196 055 |
| 1975 | 84 477 | 112 149 | 1989 | 158 576 | 202 477 |
| 1976 | 86 038 | 116 078 | 1990 | 169 238 | 223 225 |
| 1977 | 96 275 | 122 040 | 1991 | 179 001 | 233 231 |
| 1978 | 101 292 | 128 578 | 1992 | 183 687 | 242 762 |
| 1979 | 105 448 | 136 851 | 1993 | 198 273 | 259 555 |
| 1980 | 114 570 | 144 734 | | | |

*Notas:* GACP = gasto de consumo privado.
PIB = producto interno bruto.

[48] Los datos provienen del disco incluido en la obra de Chandan Mukherjee, Howard White y Marc Wuyts *Econometrics and Data Analysis for Developing Countries*, Routledge, Nueva York, 1998. Los datos originales son de las World Tables (tablas mundiales), del Banco Mundial.

**EJEMPLO 17.10** (*continuación*)

nado (17.12.3). El modelo (17.12.5) es la función de consumo de largo plazo, o de equilibrio, mientras que (17.12.6) es la función de consumo de corto plazo. $\beta_2$ mide la PMC de largo plazo, mientras que $\alpha_2$ $(= \delta\beta_2)$ da la PMC de corto plazo; la primera se obtiene de la segunda al dividir esta última entre $\delta$, el coeficiente de ajuste.

De regreso a (17.12.4), ahora podemos interpretar a 0.4043 como la PMC de corto plazo. Como $\delta = 0.4991$, la PMC de largo plazo es 0.81. Observe que el coeficiente de ajuste de alrededor de 0.50 sugiere que en cualquier periodo dado, los consumidores sólo ajustan su consumo un medio hacia su nivel deseado, o de largo plazo.

Este ejemplo permite ver el punto crucial de que, en apariencia, los modelos de expectativas adaptativas y de ajuste parcial, o el modelo de Koyck para este caso, son tan similares que no podemos decir cuál es la especificación correcta sólo con observar la regresión estimada, como (17.12.4). Por esta razón, es vital especificar la base teórica del modelo seleccionado para el análisis empírico y proceder luego apropiadamente. Si el hábito o la inercia caracterizan el comportamiento del consumo, el modelo de ajuste parcial es el adecuado. Por otra parte, si el comportamiento del consumo mira hacia adelante en el sentido de que se basa en el ingreso futuro esperado, entonces el modelo de expectativas adaptativas es el apropiado. Si es el último, se tendrá que prestar mucha atención al problema de estimación para obtener estimadores consistentes. En el primer caso, MCO proporciona estimadores consistentes en tanto se cumplan los supuestos usuales de MCO.

## 17.13 El método de Almon para los modelos de rezagos distribuidos: rezagos distribuidos polinomiales (RDP) o de Almon[49]

Si bien es muy común en la práctica, el modelo de rezagos distribuidos de Koyck se basa en el supuesto de que los coeficientes $\beta$ se reducen geométricamente a medida que aumenta el rezago (véase la figura 17.5). Este supuesto puede ser muy restrictivo en algunas situaciones. Considere, por ejemplo, la figura 17.7.

En la figura 17.7*a* se supone que las $\beta$ aumentan al principio y luego disminuyen, mientras que en la figura 17.7*c* se supone que siguen un patrón cíclico. Claro, el esquema de Koyck de modelos de rezagos distribuidos no funciona en estos casos. Sin embargo, después de mirar las figuras 17.7*a*) y *c*), parece posible expresar $\beta_i$ como función de $i$, la duración del rezago (tiempo), y ajustar curvas apropiadas para reflejar la relación funcional entre ambas, como indican las figuras 17.7*b*) y *d*). Este método es precisamente el sugerido por Shirley Almon. Para ilustrar su técnica, consideremos de nuevo el modelo de rezagos distribuidos finito ya analizado, a saber,

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \qquad \textbf{(17.1.2)}$$

que se escribe en forma más compacta como

$$Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t \qquad \textbf{(17.13.1)}$$

Según un teorema en matemáticas conocido como **teorema de Weierstrass**, Almon supone que $\beta_i$ puede aproximarse mediante un polinomio apropiado en $i$, la longitud del rezago.[50] Por ejemplo, si aplica el esquema de rezagos de la figura 17.7*a*, escribimos

$$\beta_i = a_0 + a_1 i + a_2 i^2 \qquad \textbf{(17.13.2)}$$

[49] Shirley Almon, "The Distributed Lag Between Capital Appropriations and Expenditures", *Econometrica*, vol. 33, enero de 1965, pp. 178-196.

[50] En términos generales, el teorema plantea que en un intervalo cerrado finito, cualquier función continua puede aproximarse uniformemente mediante un polinomio de un grado apropiado.

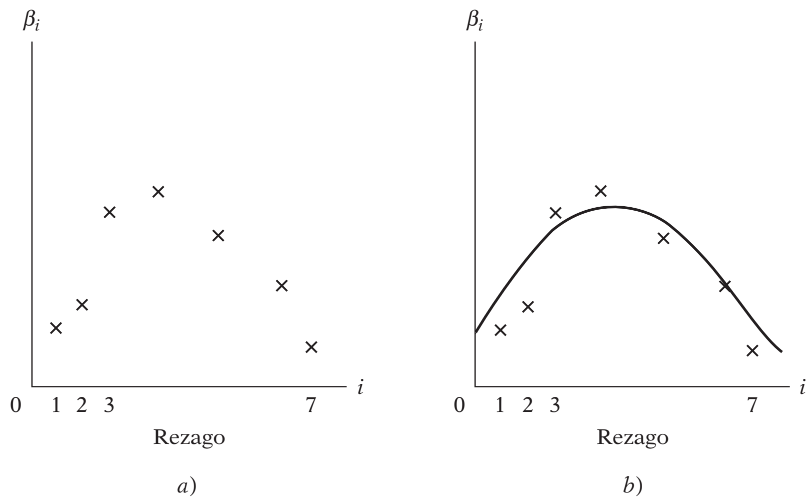



**FIGURA 17.7**
Esquema de Almon para el rezago polinomial.

que es un polinomio cuadrático, o de segundo grado, en $i$ (véase la figura 17.7*b*). Sin embargo, si las $\beta$ siguen el patrón de la figura 17.7*c*), escribimos

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 \tag{17.13.3}$$

que es un polinomio de tercer grado en $i$ (véase la figura 17.7*d*). De forma más general, podemos escribir

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m \tag{17.13.4}$$

que es un polinomio de grado $m$ en $i$. Se supone que $m$ (el grado del polinomio) es menor que $k$ (longitud máxima del rezago).

Para explicar cómo funciona el esquema de Almon, supongamos que las $\beta$ siguen el patrón de la figura 17.7*a* y, por consiguiente, la aproximación polinomial de segundo grado es apropiada. Al sustituir (17.13.2) en (17.13.1), obtenemos

$$\begin{aligned} Y_t &= \alpha + \sum_{i=0}^{k} (a_0 + a_1 i + a_2 i^2) X_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^{k} X_{t-i} + a_1 \sum_{i=0}^{k} i X_{t-i} + a_2 \sum_{i=0}^{k} i^2 X_{t-i} + u_t \end{aligned} \tag{17.13.5}$$

Definimos

$$
\begin{aligned}
Z_{0t} &= \sum_{i=0}^{k} X_{t-i} \\
Z_{1t} &= \sum_{i=0}^{k} i X_{t-i} \\
Z_{2t} &= \sum_{i=0}^{k} i^2 X_{t-i}
\end{aligned}
\tag{17.13.6}
$$

y podemos escribir (17.13.5) como

$$
Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t \tag{17.13.7}
$$

En el esquema de Almon se hace una regresión de $Y$ sobre las variables $Z$ construidas, no sobre las variables $X$ originales. Observe que (17.13.7) se estima mediante el procedimiento usual de MCO. Las estimaciones de $\alpha$ y $a_i$ así obtenidas tendrán todas las propiedades estadísticas deseables siempre que el término de perturbación estocástico $u$ satisfaga los supuestos del modelo clásico de regresión lineal. Al respecto, la técnica de Almon tiene una clara ventaja sobre el método de Koyck porque, como vimos, el último tiene algunos problemas graves de estimación debidos a la presencia de la variable explicativa estocástica $Y_{t-1}$ y a su probable correlación con el término de perturbación.

Una vez estimadas las $a$ de (17.13.7), se estiman las $\beta$ originales de (17.13.2) [o, más generalmente, de (17.13.4)] de la siguiente manera:

$$
\begin{aligned}
\hat{\beta}_0 &= \hat{a}_0 \\
\hat{\beta}_1 &= \hat{a}_0 + \hat{a}_1 + \hat{a}_2 \\
\hat{\beta}_2 &= \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 \\
\hat{\beta}_3 &= \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 \\
&\cdots\cdots\cdots\cdots\cdots \\
\hat{\beta}_k &= \hat{a}_0 + k\hat{a}_1 + k^2\hat{a}_2
\end{aligned}
\tag{17.13.8}
$$

Antes de aplicar la técnica de Almon, debemos resolver los siguientes problemas prácticos.

1. La longitud máxima del rezago $k$ debe especificarse por adelantado. Aquí quizá se puede seguir el consejo de Davidson y MacKinnon:

> El mejor enfoque es tal vez resolver primero la pregunta de la longitud del rezago, empezando con un valor muy grande de $q$ [la longitud del rezago] y luego ver si el ajuste del modelo se deteriora significativamente cuando éste se reduce sin imponer restricción alguna sobre la forma del rezago distribuido.[51]

Recuerde que si hay alguna longitud de rezago "verdadera", la selección de una cantidad menor de rezagos generará un "sesgo por omisión de variable relevante", cuyas consecuencias, como vimos en el capítulo 13, llegan a ser muy graves. Por otra parte, la selección de más rezagos de los necesarios resultará en un "sesgo por inclusión de variable irrelevante", cuyas consecuencias son menos graves; los coeficientes se estiman consistentemente por MCO, aunque sus varianzas pueden ser menos eficientes.

---

[51] Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993, pp. 675-676.

Se puede elegir la longitud adecuada del rezago con el *criterio de información de Akaike* o el de *Schwarz,* analizados en el capítulo 13. Con dichos criterios también se analiza el grado adecuado del polinomio, además de lo expuesto en el punto 2.

2. Una vez especificado $k$, debemos puntualizar también el grado $m$ del polinomio. Por lo general, el grado del polinomio debe ser por lo menos uno más que el número de puntos de inflexión en la curva que relaciona $\beta_i$ con $i$. Así, en la figura 17.7*a* sólo hay un punto de inflexión: por tanto, un polinomio de segundo grado será una buena aproximación. En la figura 17.7*c* hay dos puntos de inflexión: por tanto, un polinomio de tercer grado será una buena aproximación. *A priori*, sin embargo, no podemos conocer el número de puntos de inflexión y, por consiguiente, la selección de $m$ es muy subjetiva. Sin embargo, la teoría puede sugerir una forma particular en algunos casos. En la práctica, se espera que un polinomio de relativamente pocos grados (por ejemplo, $m = 2$ o 3) dé buenos resultados. Tras seleccionar un valor particular de $m$, si deseamos averiguar si un polinomio de mayor grado dará un mejor ajuste, procedemos de la siguiente manera.

Suponga que debemos decidir entre polinomios de segundo y tercer grados. Para el polinomio de segundo grado, la ecuación de estimación es como (17.13.7). Para el polinomio de tercer grado, la ecuación correspondiente es

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + a_3 Z_{3t} + u_t \qquad \textbf{(17.13.9)}$$

donde $Z_{3t} = \sum_{i=0}^{k} i^3 X_{t-i}$. Después de efectuar la regresión (17.13.9), si encontramos que $a_2$ es estadísticamente significativo pero que $a_3$ no lo es, podemos suponer que el polinomio de segundo grado proporciona una aproximación razonablemente buena.

Por otro lado, como proponen Davidson y MacKinnon: “Después de determinar $q$ [la longitud del rezago], se puede tratar de determinar $d$ [el grado del polinomio] empezando de nuevo con un valor grande para luego reducirlo.”[52]

Sin embargo, debemos tener cuidado con el problema de la multicolinealidad, que tal vez surgirá debido a la forma de construir las $Z$ a partir de las $X$, como se muestra en (17.13.6) [véase también (17.13.10)]. Como vimos en el capítulo 10, en casos de multicolinealidad grave, $\hat{a}_3$ puede resultar estadísticamente no significativo no porque el verdadero $a_3$ sea cero, sino tan sólo porque la muestra disponible no permite evaluar el impacto separado de $Z_3$ sobre $Y$. Por consiguiente, en el ejemplo, antes de aceptar la conclusión de que el polinomio de tercer grado no es la selección correcta, debemos asegurarnos de que el problema de multicolinealidad no sea lo bastante grave, para lo cual aplicamos las técnicas analizadas en el capítulo 10.

3. Una vez especificadas $m$ y $k$, las $Z$ se construyen fácilmente. Por ejemplo, si $m = 2$ y $k = 5$, las $Z$ son

$$\begin{aligned}
Z_{0t} &= \sum_{i=0}^{5} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5}) \\
Z_{1t} &= \sum_{i=0}^{5} i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5}) \qquad \textbf{(17.13.10)} \\
Z_{2t} &= \sum_{i=0}^{5} i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5})
\end{aligned}$$

Observe que las $Z$ son combinaciones lineales de las $X$ originales. Observe también la razón por la cual es probable que las $Z$ presenten multicolinealidad.

[52] *Ibid.*, pp. 675-676.

Antes de proceder a un ejemplo numérico, advierta las ventajas del método de Almon: en primer lugar, proporciona un método flexible de incorporar una diversidad de estructuras del rezago (véase el ejercicio 17.17). La técnica de Koyck, por otra parte, es muy rígida en el sentido de que supone que las $\beta$ se reducen geométricamente. En segundo lugar, a diferencia de la técnica de Koyck, en el método de Almon no hay que preocuparnos por la presencia de la variable dependiente rezagada como variable explicativa en el modelo y los problemas de estimación que esto crea. Por último, si se puede ajustar un polinomio de un grado lo bastante bajo, el número de coeficientes por estimar (las $a$) es mucho menor que el número original de coeficientes (las $\beta$).

Pero destaquemos de nuevo los problemas de la técnica de Almon. En primer lugar, el grado del polinomio, al igual que el valor máximo del rezago, es en gran medida una decisión subjetiva. En segundo lugar, por las razones ya mencionadas, es probable que las variables $Z$ presenten multicolinealidad. Por consiguiente, en modelos como (17.13.9) es probable que las $a$ estimadas muestren errores estándar grandes (relativos a los valores de estos coeficientes), con lo cual se obtienen uno o más de tales coeficientes estadísticamente no significativos con base en la prueba $t$ convencional. Pero esto no necesariamente significa que uno o más de los coeficientes $\hat{\beta}$ originales también sean estadísticamente no significativos. (La prueba de esta afirmación es un poco complicada, pero se sugiere en el ejercicio 17.18.) Como resultado, el problema de multicolinealidad puede no ser tan grave como se podría pensar. Además, como sabemos, en casos de multicolinealidad, aunque no sea posible estimar un coeficiente individual en forma precisa, sí lo es estimar una combinación lineal (la **función estimable**) de tales coeficientes en forma más exacta.

**EJEMPLO 17.11**
*Ilustración del modelo de rezagos distribuidos de Almon*

Para mostrar la técnica de Almon, en la tabla 17.8 se presenta información sobre inventarios $Y$ y ventas $X$ en Estados Unidos de 1954 a 1999.

Para fines ilustrativos, suponga que los inventarios dependen de las ventas en el año en curso y los tres años anteriores, de la siguiente manera:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t \quad \textbf{(17.13.11)}$$

Además, suponga que $\beta_i$ puede aproximarse mediante un polinomio de segundo grado, como en (17.13.2). Entonces, según (17.13.7), podemos escribir

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t \quad \textbf{(17.13.12)}$$

donde

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^{3} X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3}) \\ Z_{1t} &= \sum_{i=0}^{3} i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3}) \\ Z_{2t} &= \sum_{i=0}^{3} i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3}) \end{aligned} \quad \textbf{(17.13.13)}$$

Las variables $Z$ así construidas se muestran en la tabla 17.8. Con la información sobre $Y$ y las $Z$, obtenemos la siguiente regresión:

$$\begin{aligned} \hat{Y}_t &= 25\,845.06 \quad + 1.1149 Z_{0t} \quad - 0.3713 Z_{1t} \quad - 0.0600 Z_{2t} \\ \text{ee} &= (6\,596.998) \quad (0.5381) \quad (1.3743) \quad (0.4549) \\ t &= (3.9177) \quad (2.0718) \quad (-0.2702) \quad (-0.1319) \\ & R^2 = 0.9755 \quad d = 0.1643 \quad F = 517.7656 \end{aligned} \quad \textbf{(17.13.14)}$$

*Nota:* Como utilizamos un rezago de tres años, el número total de observaciones se redujo de 46 a 43.

(*continúa*)

**EJEMPLO 17.11** (*continuación*)

**TABLA 17.8** Inventarios *Y* y ventas *X*, sector manufacturero de Estados Unidos, y Z construidas

| Observación | Inventario | Ventas | $Z_0$ | $Z_1$ | $Z_2$ |
|---|---|---|---|---|---|
| 1954 | 41 612 | 23 355 | NA | NA | NA |
| 1955 | 45 069 | 26 480 | NA | NA | NA |
| 1956 | 50 642 | 27 740 | NA | NA | NA |
| 1957 | 51 871 | 28 736 | 106 311 | 150 765 | 343 855 |
| 1958 | 50 203 | 27 248 | 110 204 | 163 656 | 378 016 |
| 1959 | 52 913 | 30 286 | 114 010 | 167 940 | 391 852 |
| 1960 | 53 786 | 30 878 | 117 148 | 170 990 | 397 902 |
| 1961 | 54 871 | 30 922 | 119 334 | 173 194 | 397 254 |
| 1962 | 58 172 | 33 358 | 125 444 | 183 536 | 427 008 |
| 1963 | 60 029 | 35 058 | 130 216 | 187 836 | 434 948 |
| 1964 | 63 410 | 37 331 | 136 669 | 194 540 | 446 788 |
| 1965 | 68 207 | 40 995 | 146 742 | 207 521 | 477 785 |
| 1966 | 77 986 | 44 870 | 158 254 | 220 831 | 505 841 |
| 1967 | 84 646 | 46 486 | 169 682 | 238 853 | 544 829 |
| 1968 | 90 560 | 50 229 | 182 580 | 259 211 | 594 921 |
| 1969 | 98 145 | 53 501 | 195 086 | 277 811 | 640 003 |
| 1970 | 101 599 | 52 805 | 203 021 | 293 417 | 672 791 |
| 1971 | 102 567 | 55 906 | 212 441 | 310 494 | 718 870 |
| 1972 | 108 121 | 63 027 | 225 239 | 322 019 | 748 635 |
| 1973 | 124 499 | 72 931 | 244 669 | 333 254 | 761 896 |
| 1974 | 157 625 | 84 790 | 276 654 | 366 703 | 828 193 |
| 1975 | 159 708 | 86 589 | 307 337 | 419 733 | 943 757 |
| 1976 | 174 636 | 98 797 | 343 107 | 474 962 | 1 082 128 |
| 1977 | 188 378 | 113 201 | 383 377 | 526 345 | 1 208 263 |
| 1978 | 211 691 | 126 905 | 425 492 | 570 562 | 1 287 690 |
| 1979 | 242 157 | 143 936 | 482 839 | 649 698 | 1 468 882 |
| 1980 | 265 215 | 154 391 | 538 433 | 737 349 | 1 670 365 |
| 1981 | 283 413 | 168 129 | 593 361 | 822 978 | 1 872 280 |
| 1982 | 311 852 | 163 351 | 629 807 | 908 719 | 2 081 117 |
| 1983 | 312 379 | 172 547 | 658 418 | 962 782 | 2 225 386 |
| 1984 | 339 516 | 190 682 | 694 709 | 1 003 636 | 2 339 112 |
| 1985 | 334 749 | 194 538 | 721 118 | 1 025 829 | 2 351 029 |
| 1986 | 322 654 | 194 657 | 752 424 | 1 093 543 | 2 510 189 |
| 1987 | 338 109 | 206 326 | 786 203 | 1 155 779 | 2 688 947 |
| 1988 | 369 374 | 224 619 | 820 140 | 1 179 254 | 2 735 796 |
| 1989 | 391 212 | 236 698 | 862 300 | 1 221 242 | 2 801 836 |
| 1990 | 405 073 | 242 686 | 910 329 | 1 304 914 | 2 992 108 |
| 1991 | 390 905 | 239 847 | 943 850 | 1 389 939 | 3 211 049 |
| 1992 | 382 510 | 250 394 | 969 625 | 1 435 313 | 3 340 873 |
| 1993 | 384 039 | 260 635 | 993 562 | 1 458 146 | 3 393 956 |
| 1994 | 404 877 | 279 002 | 1 029 878 | 1 480 964 | 3 420 834 |
| 1995 | 430 985 | 299 555 | 1 089 586 | 1 551 454 | 3 575 088 |
| 1996 | 436 729 | 309 622 | 1 148 814 | 1 639 464 | 3 761 278 |
| 1997 | 456 133 | 327 452 | 1 215 631 | 1 745 738 | 4 018 860 |
| 1998 | 466 798 | 337 687 | 1 274 316 | 1 845 361 | 4 261 935 |
| 1999 | 470 377 | 354 961 | 1 329 722 | 1 921 457 | 4 434 093 |

*Nota: Y* y *X* están en millones de dólares, ajustados por estacionalidad.

Fuente: *Economic Report of the President*, 2001, tabla B-57, p. 340. Las *Z* son como las de (17.13.13).

**EJEMPLO 17.11** (*continuación*)

Cabe anotar un breve comentario sobre los resultados anteriores. De las tres variables $Z$, sólo $Z_0$ es estadísticamente significativa en un nivel individual, en el nivel de 5%; pero las otras no lo son, a pesar de que el valor $F$ es tan alto que se puede rechazar la hipótesis nula de que, de manera colectiva, las $Z$ no tienen efecto alguno sobre $Y$. Como se puede sospechar, lo anterior bien puede deberse a la multicolinealidad. Asimismo, observe que el valor $d$ calculado es muy bajo, lo cual no necesariamente significa que los residuos presenten autocorrelación. Lo más probable que el bajo valor de $d$ sugiera que el modelo esté mal especificado. Comentaremos esto más adelante.

De los coeficientes $a$ estimados dados en la ecuación (17.13.13), fácilmente se estiman los coeficientes originales $\beta$ de la relación (17.13.8). En este ejemplo, los resultados son los siguientes:

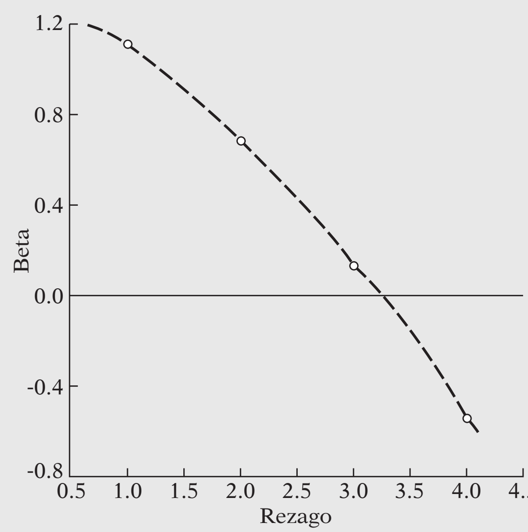
$$\begin{aligned}
\hat{\beta}_0 &= \hat{a}_0 = 1.1149 \\
\hat{\beta}_1 &= (\hat{a}_0 + \hat{a}_1 + \hat{a}_2) = 0.6836 \\
\hat{\beta}_2 &= (\hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2) = 0.1321 \\
\hat{\beta}_3 &= (\hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2) = -0.5394
\end{aligned} \tag{17.13.15}$$

Así, el modelo de rezagos distribuidos estimado correspondiente a (17.13.11) es

$$\begin{aligned}
\hat{Y}_t &= 25\,845.0 \quad + 1.1150X_0 + 0.6836X_{t-1} + 0.1321X_{t-2} - 0.5394X_{t-3} \\
\text{ee} &= (6\,596.99) \quad (0.5381) \quad (0.4672) \quad (0.4656) \quad (0.5656) \\
t &= \quad (3.9177) \quad (2.0718) \quad (1.4630) \quad (0.2837) \quad (-0.9537)
\end{aligned} \tag{17.13.16}$$

Geométricamente, las $\beta_i$ estimadas se muestran en la figura 17.8.

**FIGURA 17.8**
Estructura de rezagos del ejemplo ilustrativo.




Este ejemplo numérico sirve para señalar otras características del procedimiento de Almon:

1. Los errores estándar de los coeficientes $a$ se obtienen directamente de la regresión por MCO (17.13.14), pero los errores estándar de algunos coeficientes $\hat{\beta}$, que son el objetivo principal, no pueden obtenerse así. Pero se pueden calcular con facilidad de los errores estándar de los coeficientes $a$ estimados mediante una fórmula de estadística bien conocida, dada en el ejercicio 17.18. Desde luego, no hay necesidad de hacer esto de forma manual, pues la mayoría del software estadístico lo realiza de manera rutinaria. Los errores estándar de (17.13.15) se obtuvieron mediante *EViews 6*.

2. Las $\hat{\beta}$ obtenidas en (17.13.16) se denominan *estimaciones no restringidas*, pues no se colocan restricciones *a priori* sobre ellas. Algunas ocasiones, sin embargo, podemos desear imponer las llamadas **restricciones de punto final** sobre las $\beta$ según el supuesto de que $\beta_0$ y $\beta_k$ (el coeficiente del periodo actual y el del $k$-ésimo rezago) son cero. Por razones psicológicas, institucionales o tecnológicas, el valor de la variable explicativa en el periodo actual puede no tener impacto alguno sobre el valor actual de la variable dependiente, lo cual justifica el valor de cero para $\beta_0$. Por las mismas razones, el coeficiente del $k$-ésimo rezago puede no tener impacto alguno sobre la variable dependiente, lo cual apoya el supuesto de que $\beta_k$ es cero. En el ejemplo del inventario (ejemplo 17.11), el coeficiente de $X_{t-3}$ tuvo un signo negativo, lo cual quizá no tenga sentido económico. Por tanto, tal vez deseemos restringir ese coeficiente a cero.[53] Desde luego, no se tienen que restringir ambos extremos; podemos imponer limitantes sólo al primer coeficiente (lo cual se conoce como restricción en el extremo próximo) o al último coeficiente (restricción en el extremo lejano). Para el ejemplo del inventario, esto se ilustra en el ejercicio 17.28. A veces, las $\beta$ están estimadas con la restricción de que su suma sea igual a la unidad. Pero no se deben imponer restricciones de manera negligente, pues también afectan los valores de los demás coeficientes rezagados (sin restricciones).
3. En vista de que la elección del número de coeficientes rezagados y el grado del polinomio se dejan al criterio del que elabora el modelo, resulta inevitable cierto grado de ensayo y error, además de la carga que implica la búsqueda exhaustiva de datos. En esta situación es cuando resultan útiles los **criterios de información de Akaike y Schwarz**, analizados en el capítulo 13.
4. Como estimamos (17.13.16) mediante tres rezagos y el polinomio de segundo grado, es un modelo de mínimos cuadrados restringidos. Suponga que decidimos usar los tres rezagos pero sin el método polinomial de Almon; es decir, calculamos (17.13.11) mediante MCO. ¿Qué sucede entonces? Primero veamos los resultados:

$$
\begin{aligned}
\hat{Y}_t &= 26\,008.60 \quad + 0.9771X_t \quad + 1.0139X_{t-1} - \quad 0.2022\,X_{t-2} - \quad 0.3935X_{t-3}\\
\text{ee} &= (6\,691.12) \quad (0.6820) \quad (1.0920) \quad (1.1021) \quad (0.7186)\\
t &= \quad (3.8870) \quad (1.4327) \quad (0.9284) \quad (-0.1835) \quad (-0.5476)\\
&\qquad R^2 = 0.9755 \qquad d = 0.1571 \qquad F = 379.51
\end{aligned}
\tag{17.13.17}
$$

Si compara estos resultados con los de (17.13.16), verá que la $R^2$ general es prácticamente la misma, aunque el patrón de rezago en (17.13.17) muestra una forma más corvada que la de (17.13.16). Se deja al lector que verifique el valor de $R^2$ de (17.13.16).

Como ilustra este ejemplo, se debe tener cuidado con la técnica de rezagos distribuidos de Almon, pues los resultados pueden llegar a ser sensibles a la hora de elegir el grado del polinomio y/o el número de coeficientes rezagados.

## 17.14 Causalidad en economía: prueba de causalidad de Granger[54]

Antes, en la sección 1.4, mencionamos que, aunque el análisis de regresión trata sobre la dependencia de una variable sobre otras variables, esto no necesariamente implica causalidad. En otras palabras, la existencia de una relación entre las variables no prueba causalidad ni la dirección de

[53] Para una aplicación concreta, véase D.B. Batten y Daniel Thornton, "Polynomial Distributed Lags and the Estimation of the St. Louis Equation", *Review*, Federal Reserve Bank of St. Louis, abril de 1983, pp. 13-25.

[54] Existe otra prueba de causalidad que a veces se utiliza, la llamada **prueba de causalidad de Sims**. La estudiaremos mediante un ejercicio.

la influencia. Pero en regresiones con datos de series de tiempo, la situación puede ser un cuanto distinta porque, como sostiene un autor:

> . . . el tiempo no corre hacia atrás. Es decir, si un acontecimiento *A* sucede antes de un suceso *B*, es *posible* que *A* cause *B*. Sin embargo, no es *posible* que *B* provoque *A*. En otras palabras, los acontecimientos pasados pueden propiciar sucesos que ocurren en la actualidad. Lo cual no sucede con los sucesos futuros.[55] (Cursivas nuestras.)

Ésta es la idea general de la llamada prueba de causalidad de Granger.[56] Pero debemos notar con toda claridad que el asunto de la causalidad es en gran medida filosófico, con toda la polémica que esto conlleva. En un extremo están quienes opinan que "todo es causa de todo", y en el otro, quien niega la existencia de cualquier clase de causalidad.[57] El econometrista Edward Leamer prefiere el término **precedencia** en vez de causalidad. Francis Diebold se inclina más por el término **causalidad predictiva**:

> . . . el enunciado "$y_i$ causa $y_j$" es sólo una forma abreviada del enunciado más preciso aunque más extenso de que "$y_i$ contiene información útil para predecir $y_j$ (en el sentido de los mínimos cuadrados lineales), además de las historias pasadas de las demás variables del sistema". Para ahorrar espacio, sólo decimos que $y_i$ causa $y_j$.[58]

## Prueba de Granger

Para explicar la prueba de Granger consideraremos una pregunta frecuente en macroeconomía: ¿el PIB "causa" la oferta de dinero *M* (PIB → *M*), o la oferta de dinero *M* causa el PIB (*M* → PIB)? (La flecha señala la dirección de la causalidad.) La prueba de causalidad de Granger supone que la información relevante para la predicción de las variables respectivas, PIB y *M*, está contenida únicamente en la información de series de tiempo sobre estas variables. La prueba implica la estimación de las siguientes regresiones:

$$\text{PIB}_t = \sum_{i=1}^{n} \alpha_i M_{t-i} + \sum_{j=1}^{n} \beta_j \text{PIB}_{t-j} + u_{1t} \tag{17.14.1}$$

$$M_t = \sum_{i=1}^{n} \lambda_i M_{t-i} + \sum_{j=1}^{n} \delta_j \text{PIB}_{t-j} + u_{2t} \tag{17.14.2}$$

donde se supone que las perturbaciones $u_{1t}$ y $u_{2t}$ no están correlacionadas. A propósito, observe que en vista de que hay dos variables, tratamos con una **causalidad bilateral**. En los capítulos sobre econometría de series de tiempo, la generalizaremos a la causalidad multivariada mediante la técnica de **autorregresión vectorial (VAR)**.

La ecuación (17.14.1) postula que el PIB actual se relaciona con los valores pasados del PIB mismo, al igual que con los de *M*, y (17.14.2) postula un comportamiento similar para $M_t$.

---

[55] Gary Koop, *Analysis of Economic Data*, John Wiley & Sons, Nueva York, 2000, p. 175.

[56] C.W.J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, julio de 1969, pp. 424-438. Aunque se conoce más como prueba Granger de causalidad, resulta apropiado llamarla **prueba de causalidad de Wiener-Granger**, en vista de que antes la propuso Wiener. Véase N. Wiener, "The Theory of Prediction", en E.F. Beckenback (ed.), *Modern Mathematics for Engineers*, McGraw-Hill, Nueva York, 1956, pp. 165-190.

[57] Hay un excelente análisis de este tema en Arnold Zellner, "Causality and Econometrics", *Carnegie-Rochester Conference Series*, 10, K. Brunner y A.H. Meltzer (eds.), North Holland, Ámsterdam, 1979, pp. 9-50.

[58] Francis X. Diebold, *Elements of Forecasting*, South Western, 2a. ed., 2001, p. 254.

Observe que estas regresiones se realizan en forma de crecimientos, PIB y $M$, donde un punto sobre una variable indica su tasa de crecimiento. Ahora distinguimos cuatro casos:

1. La *causalidad unidireccional de M hacia PIB* es la indicada si los coeficientes estimados sobre la $M$ rezagada en (17.14.1) son estadísticamente diferentes de cero considerados como grupo y el conjunto de coeficientes estimados sobre el PIB rezagado en (17.14.2) no es estadísticamente diferente de cero.
2. Al contrario, hay *causalidad unidireccional de PIB hacia M* si el conjunto de coeficientes de $M$ rezagada en (17.14.1) no es estadísticamente diferente de cero y el conjunto de coeficientes del PIB rezagado en (17.14.2) es estadísticamente diferente de cero.
3. Se sugiere *realimentación*, o *causalidad bilateral*, cuando los conjuntos de coeficientes de $M$ y de PIB son estadísticamente significativos, diferentes de cero, en ambas regresiones.
4. Por último, se sugiere *independencia* cuando los conjuntos de coeficientes de $M$ y de PIB no son estadísticamente significativos en ambas regresiones.

De modo más general, como el futuro no puede predecir el pasado, si la variable $X$ (a la manera de Granger) causa la variable $Y$, los cambios en $X$ deben *preceder* a los cambios en $Y$. Por consiguiente, en una regresión de $Y$ sobre otras variables (con sus propios valores pasados), si incluimos valores pasados o rezagados de $X$ y esto mejora significativamente la predicción de $Y$, podemos decir que $X$ (a la manera de Granger) causa $Y$. Aplica una definición similar si $Y$ (a la manera de Granger) causa $X$.

Los pasos comprendidos en la prueba de causalidad de Granger son los siguientes. Ilustramos estos pasos con el ejemplo PIB-dinero de la ecuación (17.14.1).

1. Haga la regresión del PIB actual sobre todos los términos rezagados del PIB y otras variables, en su caso, pero *no* incluya las variables $M$ rezagadas en esta regresión. De acuerdo con el capítulo 8, ésta es la regresión restringida. A partir de esta regresión, obtenga la suma de cuadrados residuales restringida, $\text{SCR}_R$.
2. Ahora efectúe la regresión con los términos rezagados $M$. En el lenguaje del capítulo 8, ésta es la regresión no restringida. A partir de esta regresión, obtenga la suma de cuadrados residuales no restringida, $\text{SCR}_{\text{NR}}$.
3. La hipótesis nula es $H_0$: $\alpha_i = 0$, $i = 1, 2, \ldots, n$, es decir, los términos rezagados de $M$ no pertenecen a la regresión.
4. Para probar esta hipótesis, aplicamos la prueba $F$ dada por (8.7.9), a saber,

$$F = \frac{(\text{SCR}_R - \text{SCR}_{\text{NR}})/m}{\text{SCR}_{\text{NR}}/(n-k)} \tag{8.7.9}$$

   que sigue la distribución $F$ con $m$ y $(n - k)$ gl. En el presente caso, $m$ es igual al número de términos rezagados de $M$, y $k$ es el número de parámetros estimados en la regresión no restringida.
5. Si el valor $F$ calculado excede al valor $F$ crítico en el nivel seleccionado de significancia, rechazamos la hipótesis nula, en cuyo caso los términos rezagados de $M$ pertenecen a la regresión. Ésta es otra forma de decir que $M$ causa al PIB.
6. Se repiten los pasos 1 a 5 para probar el modelo (17.14.2), es decir, para definir si el PIB causa $M$.

Antes de ilustrar la prueba de causalidad de Granger, hay varios puntos que deben tomarse en cuenta:

1. Se supone que las dos variables, PIB y $M$, son *estacionarias*. Ya estudiamos el concepto de estacionariedad en términos intuitivos y lo analizaremos de manera más formal en el capítulo 21. A veces, al tomar las primeras diferencias de las variables, las convierte en estacionarias, si es que aún no son estacionarias en la forma de nivel.

2. El número de términos rezagados por introducir en las pruebas de causalidad es una cuestión práctica importante. Como en el caso de los modelos de rezagos distribuidos, quizá tengamos que elegir mediante el criterio de información Akaike o Schwarz. Pero debe añadirse que la *dirección de la causalidad tal vez dependa de manera crítica del número de términos rezagados incluidos*.
3. Supusimos que los términos de error en la prueba de causalidad no están correlacionados. De no ser así, quizá deba llevarse a cabo la transformación adecuada, como analizamos en el capítulo 12.[59]
4. Como nos interesa probar la causalidad, no se requiere presentar los coeficientes estimados de los modelos (17.14.1) y (17.14.2) en forma explícita (para ahorrar espacio); bastan los resultados de la prueba *F* dados en (8.7.9).
5. Es necesario tener cuidado con la causalidad "espuria". En el ejemplo del PIB y la oferta de dinero, suponga que consideramos la tasa de interés (por ejemplo, la de corto plazo). Es muy posible que el dinero "cause según Granger" la tasa de interés y que la tasa de interés, a su vez, "cause según Granger" el PIB. Por tanto, si no tomamos en cuenta la tasa de interés y concluimos que el dinero causa el PIB, la causalidad observada entre el PIB y la oferta de dinero puede ser espuria.[60] Como ya señalamos, una forma de resolver esto es considerar un sistema de ecuaciones múltiples, como la **autorregresión vectorial (VAR),** que estudiaremos con cierto detalle en el capítulo 22.

**EJEMPLO 17.12**
*Causalidad entre dinero e ingreso*

R. W. Hafer estableció, con la prueba de Granger, la naturaleza de la causalidad entre el PNB (en vez del PIB) y *M* para Estados Unidos de 1960-I a 1980-IV. En vez de los valores brutos de estas variables, empleó sus tasas de crecimiento, de PNB y *M*; asimismo, utilizó cuatro rezagos de cada variable en las dos regresiones dadas antes. Los resultados fueron los siguientes:[61] En cada caso, la hipótesis nula establece que la variable en consideración no causa "a la manera de Granger" la otra variable.

| Dirección de la causalidad | Valor *F* | Decisión |
|---|---|---|
| $\dot{M} \rightarrow \dot{\text{PNB}}$ | 2.68 | Se rechaza |
| $\dot{\text{PNB}} \rightarrow \dot{M}$ | 0.56 | No se rechaza |

Estos resultados sugieren que la dirección de la causalidad va del crecimiento del dinero al del PNB, pues el valor *F* estimado es significativo en el nivel de 5%; el valor *F* crítico es 2.50 (para 4 y 71 gl). Por otra parte, no hay una "causalidad inversa" del crecimiento del PNB al del dinero, pues el valor *F* calculado no es estadísticamente significativo.

**EJEMPLO 17.13**
*Causalidad entre dinero y tasa de interés en Canadá*

Consulte los datos sobre Canadá proporcionados en la tabla 17.5. Suponga que deseamos averiguar si existe causalidad entre la oferta de dinero y la tasa de interés en Canadá en periodos trimestrales de 1979 a 1988. A fin de demostrar que la prueba de Granger para causalidad depende de manera crítica del número de términos rezagados introducidos en el modelo, presentamos enseguida los resultados de la prueba *F* en la que se emplearon varios rezagos (trimestrales). En cada caso, la hipótesis nula es que la tasa de interés *no* causa (a la manera de Granger) la oferta de dinero y viceversa.

(*continúa*)

[59] Para mayores detalles, véase Wojciech W. Charemza y Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modeling, Cointegration and Vector Autoregression*, 3a. ed., Edward Elgar, 1997, cap. 6.

[60] Sobre este tema, véase J.H. Stock y M.W. Watson, "Interpreting the Evidence on Money-Income Causality", *Journal of Econometrics,* vol. 40, 1989, pp. 783-820.

[61] R.W. Hafer, "The Role of Fiscal Policy in the St. Louis Equation", *Review*, Federal Reserve Bank of St. Louis, enero de 1982, pp. 17-22. Véase la nota 12 de su obra para los detalles del procedimiento.

**EJEMPLO 17.13**
(*continuación*)

| Dirección de la causalidad | Número de rezagos | Valor *F* | Decisión |
|---|---|---|---|
| $R \rightarrow M$ | 2 | 12.92 | Rechazar |
| $M \rightarrow R$ | 2 | 3.22 | Rechazar |
| $R \rightarrow M$ | 4 | 5.59 | Rechazar |
| $M \rightarrow R$ | 4 | 2.45 | Rechazar (en 7%) |
| $R \rightarrow M$ | 6 | 3.5163 | Rechazar |
| $M \rightarrow R$ | 6 | 3.71 | Rechazar |
| $R \rightarrow M$ | 8 | 1.40 | No rechazar |
| $M \rightarrow R$ | 8 | 1.62 | No rechazar |

Observe estas características de los resultados anteriores de la prueba *F*: de uno a seis rezagos existe una causalidad bilateral entre la oferta de dinero y la tasa de interés. No obstante, a los ocho rezagos, no existe una relación estadísticamente discernible entre las dos variables. Lo anterior refuerza la observación anterior de que el resultado de la prueba de Granger es sensible al número de rezagos introducidos en el modelo.

**EJEMPLO 17.14**
*Causalidad entre la tasa de crecimiento del PIB y la tasa bruta del ahorro en nueve países de Asia oriental*

Un estudio de la causalidad bilateral entre la tasa de crecimiento del PIB ($g$) y la tasa bruta del ahorro ($s$) mostró los resultados de la tabla 17.9.[62] Con fines comparativos, los resultados también se presentan en la tabla los resultados de Estados Unidos. Por mucho, los resultados de la tabla 17.9 muestran que para la mayoría de los países de Asia oriental la causalidad tiene una dirección desde la tasa de crecimiento del PIB hacia la tasa bruta del ahorro. En contraste, para Estados Unidos la causalidad se da en ambas direcciones de 1950 a 1988 hasta el tercer rezago; pero a partir del cuarto y quinto rezagos, la causalidad va de la tasa de crecimiento del PIB hacia la tasa de ahorro, pero no en sentido inverso.

**TABLA 17.9**
**Prueba de causalidad bivariada de Granger entre la tasa de crecimiento real del PIB *per cápita* y la tasa bruta del ahorro**

Fuente: Banco Mundial, *The East Asian Miracle: Economic Growth and Public Policy*, Oxford University Press, Nueva York, 1993, p. 244 (tabla A5-2). La fuente original es Robert Summers y Alan Heston, "The Penn World Tables (Mark 5): An Expanded Set of International Comparisons, 1950-88", *Quarterly Journal of Economics*, vol. 105, núm. 2, 1991.

| Economía, años | Años de rezago | Variable del lado derecho rezagada Ahorro | Crecimiento | Economía, años | Años de rezago | Variable del lado derecho rezagada Ahorro | Crecimiento |
|---|---|---|---|---|---|---|---|
| Hong Kong, | 1 | Sig | Sig | Filipinas, | 1 | NS | Sig |
| 1960-1988 | 2 | Sig | Sig | 1950-1988 | 2 | NS | Sig |
| | 3 | Sig | Sig | | 3 | NS | Sig |
| | 4 | Sig | Sig | | 4 | NS | Sig |
| | 5 | Sig | Sig | | 5 | NS | Sig |
| Indonesia, | 1 | Sig | Sig | Singapur, | 1 | NS | NS |
| 1965 | 2 | NS | Sig | 1960-1988 | 2 | NS | NS |
| | 3 | NS | Sig | | 3 | NS | NS |
| | 4 | NS | Sig | | 4 | Sig | NS |
| | 5 | NS | Sig | | 5 | Sig | NS |
| Japón, | 1 | NS | Sig | Taiwán, China, | 1 | Sig | Sig |
| 1950-1988 | 2 | NS | Sig | 1950-1988 | 2 | NS | Sig |
| | 3 | NS | Sig | | 3 | NS | Sig |
| | 4 | NS | Sig | | 4 | NS | Sig |
| | 5 | NS | Sig | | 5 | NS | Sig |
| Corea, Rep. de, | 1 | Sig | Sig | Tailandia, | 1 | NS | Sig |
| 1955-1988 | 2 | NS | Sig | 1950-1988 | 2 | NS | Sig |
| | 3 | NS | Sig | | 3 | NS | Sig |
| | 4 | NS | Sig | | 4 | NS | Sig |
| | 5 | NS | Sig | | 5 | NS | Sig |
| Malasia, | 1 | Sig | Sig | Estados Unidos, | 1 | Sig | Sig |
| 1955-1988 | 2 | Sig | Sig | 1950-1988 | 2 | Sig | Sig |
| | 3 | NS | NS | | 3 | Sig | Sig |
| | 4 | NS | NS | | 4 | NS | Sig |
| | 5 | NS | Sig | | 5 | NS | Sig |

Sig.: significativa; NS: no significativa.
*Nota:* El crecimiento es el crecimiento real *per cápita* del PIB a precios internacionales de 1985.

[62] Estos resultados provienen de *The East Asian Miracle: Growth and Public Policy*, publicado para el Banco Mundial por Oxford University Press, 1993, p. 244.

**EJEMPLO 17.14** (*continuación*)

Para concluir el análisis de causalidad de Granger, tenga en cuenta que la pregunta es si se puede detectar *estadísticamente* la dirección de la causalidad cuando *temporalmente* existe una relación entre los valores rezagados-adelantados de dos variables. Si se establece la causalidad, ésta sugeriría que puede utilizarse una variable para predecir la otra variable mejor de lo que se haría si se usara simplemente el pasado histórico de esta última variable. En el caso de las economías del este asiático, parece que puede predecirse mejor la tasa bruta del ahorro si se consideran los valores rezagados de la tasa de crecimiento del PIB, en vez de tener en cuenta sólo los valores rezagados de la tasa bruta del ahorro.

## *Nota sobre causalidad y exogeneidad

Como estudiaremos en los capítulos sobre el modelo con ecuaciones simultáneas en la **Parte 4** de esta obra, a menudo las variables económicas se clasifican según dos amplias categorías: **endógenas** y **exógenas**. En términos generales, las variables endógenas equivalen a las variables dependientes de un modelo de regresión con una sola ecuación, y las exógenas, a las variables $X$, o regresoras, en tal modelo, en tanto las variables $X$ no estén correlacionadas con el término de error de esa ecuación.[63]

Ahora surge una pregunta importante: suponga que mediante la prueba de causalidad de Granger descubrimos que una variable $X$ causa (a la manera de Granger) una variable $Y$, sin que esta última cause a aquélla (es decir, no hay causalidad bilateral), ¿podemos considerar entonces exógena la variable $X$? En otras palabras, ¿con la causalidad de Granger (o la no causalidad) podemos establecer exogeneidad?

Para responder necesitamos distinguir tres tipos de exogeneidad: 1) débil, 2) fuerte y 3) súper. A fin de conservar la simplicidad en la exposición, suponga que sólo tenemos dos variables, $Y_t$ y $X_t$, y que hacemos la regresión de $Y_t$ sobre $X_t$. Decimos que $X_t$ es *débilmente exógena* si $Y_t$, además, no explica la existencia de $X$. En este caso, la estimación y la prueba del modelo de regresión pueden hacerse condicionales a los valores de $X_t$. En realidad, recuerde, del capítulo 2, que observamos que el modelo de regresión estaba condicionado a los valores de las variables $X$. Decimos que $X_t$ es *fuertemente exógena* si los valores $Y$, actuales y rezagados, no la explican (es decir, no existe una relación de realimentación). Asimismo, $X_t$ es *superexógena* si los parámetros en la regresión de $Y$ y $X$ no cambian aunque sí lo hagan los valores de $X$; es decir, los valores de los parámetros son invariantes ante los cambios de (los) valor(es) de $X$. Si así sucediese en efecto, la famosa "crítica de Lucas" perdería su fuerza.[64]

La razón para distinguir entre los tres tipos de exogeneidad es que "En general, la exogeneidad débil es todo lo que se requiere para la estimación y la comprobación, la exogeneidad fuerte es necesaria para la predicción, y la superexogeneidad se necesita para el análisis de políticas".[65]

De regreso con el tema de la causalidad de Granger, si una variable, por ejemplo $Y$, no causa otra variable, digamos $X$, ¿se puede suponer entonces que esta última es exógena? Por desgracia, la respuesta no es inmediata. Si hablamos de exogeneidad débil, podemos demostrar que la *causalidad de Granger no es necesaria ni suficiente para establecer la exogeneidad*. Por otra parte, la causalidad de Granger es necesaria (pero no suficiente) para la exogeneidad fuerte. Las pruebas de ambos enunciados trascienden el alcance de este libro.[66] Así, para nuestro propósito, es mejor

* Opcional.

[63] Desde luego, si las variables explicativas incluyen uno o más términos rezagados de la variable endógena, quizá no se satisfaga este requisito.

[64] Robert Lucas, ganador del Premio Nobel, planteó que las relaciones entre las variables económicas pueden cambiar cuando las políticas varían, en cuyo caso los parámetros estimados a partir de un modelo de regresión tendrán escaso valor para la predicción. Sobre este tema, véase Oliver Blanchard, *Macroeconomics*, Prentice Hall, 1997, pp. 371-372.

[65] Keith Cuthbertson, Stephen G. Hall y Mark P. Taylor, *Applied Econometric Techniques*, University of Michigan Press, 1992, p. 100.

[66] Para un análisis comparativamente sencillo, véase G.S. Maddala, *Introduction to Econometrics*, 2a. ed., Macmillan, Nueva York, 1992, pp. 394-395, y también David F. Hendry, *Dynamic Econometrics*, Oxford University Press, Nueva York, capítulo 5.

conservar separados los conceptos de causalidad de Granger y exogeneidad, y tratar al primero como una herramienta descriptiva útil para los datos de las series de tiempo. En el capítulo 19 analizaremos una prueba para averiguar si una variable puede considerarse exógena.

## Resumen y conclusiones

1. Por razones psicológicas, tecnológicas e institucionales, una variable regresada puede responder a una o varias regresoras con un rezago de tiempo. Los modelos de regresión que tienen en cuenta rezagos de tiempo se conocen como **modelos de regresión dinámicos** o **rezagados**.
2. Hay dos tipos de modelos rezagados: **de rezagos distribuidos** y **autorregresivos**. En el primero, los valores actuales y rezagados de las regresoras son variables explicativas. En el último, el (los) valor(es) rezagado(s) de la regresada aparece(n) como variable(s) explicativa(s).
3. Un modelo puramente de rezagos distribuidos se estima mediante MCO, pero en ese caso aparece el problema de multicolinealidad, pues los valores rezagados sucesivos de una regresora tienden a estar correlacionados.
4. Como resultado, se han diseñado algunos métodos abreviados: los mecanismos de Koyck, de expectativas adaptativas y de ajuste parcial. El primero es un método puramente algebraico y los otros dos se basan en principios económicos.
5. Una característica única de los modelos de **Koyck, de expectativas adaptativas** y **de ajuste parcial** es que todos son autorregresivos por naturaleza, es decir, el valor o valores de la variable regresada aparecen como una de las variables explicativas.
6. La autorregresividad plantea desafíos en la estimación; si la variable regresada rezagada está correlacionada con el término de error, los estimadores de MCO de tales modelos no sólo están sesgados, sino que también son inconsistentes. El sesgo y la inconsistencia se presentan con los modelos de Koyck y de expectativas adaptativas; el modelo de ajuste parcial es diferente y se estima consistentemente mediante MCO, a pesar de la presencia de la variable regresada rezagada.
7. Para estimar los modelos de Koyck y de expectativas adaptativas consistentemente, el método más común es el **método de variables instrumentales**. La variable instrumental es una variable representante para la variable regresada rezagada pero con la propiedad de que no está correlacionada con el término de error.
8. Una alternativa para los modelos rezagados de regresión recién analizada es el **modelo de rezagos distribuidos polinomial de Almon**, con el cual se evitan los problemas de estimación asociados a los modelos autorregresivos. El principal problema con el método de Almon, sin embargo, es que se debe *especificar por anticipado* la longitud del rezago y el grado del polinomio. Hay métodos formales e informales para resolver la selección de la longitud del rezago y el grado del polinomio.
9. A pesar de los problemas de estimación, que pueden resolverse, los modelos distribuidos y autorregresivos han demostrado ser muy útiles en la economía empírica, porque con ellos es posible dinamizar la teoría económica que de otra forma sería estática, al tener en cuenta explícitamente el papel del tiempo. Tales modelos permiten diferenciar respuestas de corto y largo plazos de la variable dependiente ante cambios unitarios en el valor de la(s) variable(s) explicativa(s). Así, para estimar los plazos corto y largo de precio, ingreso, sustitución y otras elasticidades, estos modelos han demostrado ser muy útiles.[67]
10. Debido a los rezagos participantes, los modelos distribuidos y/o autorregresivos conducen al tema de la causalidad en las variables económicas. En la práctica, la elaboración de modelos de **causalidad de Granger** ha recibido considerable atención. Pero se debe tener mucho

[67] Para aplicaciones de estos modelos, véase Arnold C. Harberger (ed.), *The Demand for Durable Goods*, University of Chicago Press, Chicago, 1960.

cuidado con la metodología de Granger, porque es muy sensible a la longitud del rezago utilizado en el modelo.

11. Aunque una variable ($X$) cause (a la manera de Granger) otra variable ($Y$), eso no significa que $X$ sea exógena. Distinguimos tres tipos de exogeneidad (débil, fuerte y súper) y señalamos la importancia de dicha clasificación.

## EJERCICIOS

### Preguntas

17.1. Explique, de manera breve, si las siguientes afirmaciones son verdaderas, falsas o inciertas:

*a*) Todos los modelos econométricos son en esencia dinámicos.

*b*) El modelo de Koyck no tiene mucho sentido si algunos coeficientes de los rezagos distribuidos son positivos y otros negativos.

*c*) Si los modelos de Koyck y de expectativas adaptativas se estiman mediante MCO, los estimadores serán sesgados pero consistentes.

*d*) En el modelo de ajuste parcial, los estimadores de MCO son sesgados en muestras finitas.

*e*) En presencia de una o varias regresoras estocásticas y de un término de error autocorrelacionado, el método de variables instrumentales produce estimaciones insesgadas y consistentes.

*f*) En presencia de una variable regresada rezagada como variable regresora, el estadístico $d$ de Durbin-Watson para detectar autocorrelación es prácticamente inútil.

*g*) La prueba $h$ de Durbin es válida en muestras grandes y pequeñas.

*h*) La prueba de Granger es una prueba de precedencia más que de causalidad.

17.2. Desarrolle la ecuación (17.7.2).

17.3. Demuestre la ecuación (17.8.3).

17.4. Suponga que los precios se forman de acuerdo con la siguiente hipótesis de expectativas adaptativas:

$$P_t^* = \gamma P_{t-1} + (1-\gamma)P_{t-1}^*$$

donde $P^*$ es el precio esperado y $P$ es el precio observado.

Complete la siguiente tabla, suponiendo que $\gamma = 0.5$:*

| Periodo | P* | P |
|---|---|---|
| $t-3$ | 100 | 110 |
| $t-2$ | | 125 |
| $t-1$ | | 155 |
| $t$ | | 185 |
| $t+1$ | | — |

17.5. Considere el modelo

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 Y_{t-1} + v_t$$

* Adaptado de G.K. Shaw, *op. cit.*, p. 26.

Suponga que $Y_{t-1}$ y $v_t$ están correlacionadas. Para eliminar la correlación, suponga que utilizamos el siguiente método de variable instrumental: primero haga la regresión de $Y_t$ sobre $X_{1t}$ y $X_{2t}$ y obtenga el estimado $\hat{Y}_t$ de esta regresión. Luego efectúe la regresión

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 \hat{Y}_{t-1} + v_t$$

donde $\hat{Y}_{t-1}$ se estiman de la regresión de la primera etapa.

*a*) ¿Cómo elimina este procedimiento la correlación entre $Y_{t-1}$ y $v_t$ en el modelo original?

*b*) ¿Cuáles son las ventajas del procedimiento recomendado sobre el enfoque de Liviatan?

*17.6. *a*) Desarrolle (17.4.8).

*b*) Evalúe el rezago medio de $\lambda = 0.2, 0.4, 0.6, 0.8$.

*c*) ¿Hay alguna relación sistemática entre el valor de λ y el rezago medio?

17.7. *a*) Pruebe que para el modelo de Koyck, el rezago medio es como el que aparece en (17.4.10).

*b*) Si λ es relativamente grande, ¿cuáles son sus implicaciones?

17.8. Con la fórmula para el rezago medio dada en (17.4.9), verifique el rezago medio de 10.959 trimestres reportado en la ilustración de la tabla 17.1.

17.9. Suponga que

$$M_t = \alpha + \beta_1 Y_t^* + \beta_2 R_t^* + u_t$$

donde $M =$ demanda de saldos reales de efectivo, $Y^* =$ ingreso real esperado y $R^* =$ tasa de interés esperada. Suponga que las expectativas se formulan de la siguiente manera:

$$Y_t^* = \gamma_1 Y_t + (1 - \gamma_1) Y_{t-1}^*$$
$$R_t^* = \gamma_2 R_t + (1 - \gamma_2) R_{t-1}^*$$

donde $\gamma_1$ y $\gamma_2$ son los coeficientes de expectativas, ambos entre 0 y 1.

*a*) ¿Cómo expresa $M_t$ en términos de las cantidades observables?

*b*) ¿Qué problemas de estimación prevé?

*17.10. Si estima (17.7.2) por MCO, ¿puede derivar estimaciones de los parámetros originales? ¿Qué problemas prevé? (Para mayores detalles, vea Roger N. Waud).†

17.11. *Modelo de correlación serial*. Considere el siguiente modelo:

$$Y_t = \alpha + \beta X_t + u_t$$

Suponga que $u_t$ sigue el esquema autorregresivo de Markov de primer orden dado en el capítulo 12, a saber,

$$u_t = \rho u_{t-1} + \varepsilon_t$$

donde $\rho$ es el coeficiente de autocorrelación (de primer orden) y donde $\varepsilon_t$ satisface todos los supuestos clásicos de MCO. Entonces, como vimos en el capítulo 12, el modelo

$$Y_t = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + \rho Y_{t-1} + \varepsilon_t$$

tendrá un término de error serialmente independiente, lo que posibilita la estimación por MCO. Pero este modelo, llamado **modelo de correlación serial**, se parece mucho a

* Opcional.

† "Misspecification in the 'Partial Adjustment' and 'Adaptive Expectations' Models", *International Economic Review*, vol. 9, núm. 2, junio de 1968, pp. 204-217.

los modelos de Koyck, de expectativas adaptativas y de ajuste parcial. ¿Cómo sabría, en cualquier situación dada, cuál de los modelos anteriores es el apropiado?*

17.12. Considere el modelo de Koyck (o, para ese caso, el de expectativas adaptativas) dado en (17.4.7), a saber,

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (u_t - \lambda u_{t-1})$$

Suponga que en el modelo original, $u_t$ sigue el esquema autorregresivo de primer orden $u_t - \rho u_{1-t} = \varepsilon_t$, donde $\rho$ es el coeficiente de autocorrelación y donde $\varepsilon_t$ satisface todos los supuestos clásicos de MCO.

*a*) Si $\rho = \lambda$, ¿el modelo de Koyck puede estimarse mediante MCO?

*b*) ¿Serán insesgadas las estimaciones así obtenidas? ¿Consistentes? ¿Por qué?

*c*) ¿Qué tan razonable es suponer que $\rho = \lambda$?

17.13. *Modelo de rezagos distribuidos triangular o aritmético.*† Este modelo supone que el estímulo (la variable explicativa) ejerce su mayor impacto en el tiempo actual y luego se reduce en cantidades iguales hasta llegar a cero a medida que uno retrocede en el pasado distante. Geométricamente, esto se muestra en la figura 17.9. Según esta distribución, suponga que efectuamos la siguiente sucesión de regresiones:

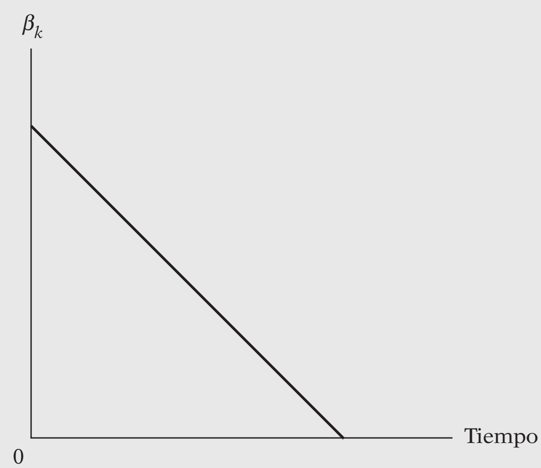
$$Y_t = \alpha + \beta\left(\frac{2X_t + X_{t-1}}{3}\right)$$

$$Y_t = \alpha + \beta\left(\frac{3X_t + 2X_{t-1} + X_{t-2}}{6}\right)$$

$$Y_t = \alpha + \beta\left(\frac{4X_t + 3X_{t-1} + 2X_{t-2} + X_{t-1}}{10}\right)$$

etc., y seleccionamos la regresión que da la $R^2$ más elevada como la "mejor" regresión. Comente esta estrategia.

**FIGURA 17.9**
Esquema de rezagos aritmético o triangular (de Fisher).




* Hay un análisis del modelo de correlación serial en Zvi Griliches, "Distributed Lags: A Survey", *Econometrica*, vol. 35, núm. 1, enero de 1967, p. 34.

† Irving Fisher propuso este modelo en "Note on a Short-Cut Method for Calculating Distributed Lags", *International Statistical Bulletin*, 1937, pp. 323-328.

17.14. Con base en la información trimestral del periodo 1950-1960, F. P. R. Brechling obtuvo la siguiente función de demanda de trabajo para la economía británica (las cifras entre paréntesis son errores estándar):*

$$\widehat{\dot{E}}_t = \underset{(2.61)}{14.22} + \underset{(0.014)}{0.172Q_t} - \underset{(0.015)}{0.028t} - \underset{(0.0002)}{0.0007t^2} - \underset{(0.033)}{0.297E_{t-1}}$$

$$\bar{R}^2 = 0.76 \qquad d = 1.37$$

donde $\dot{E}_t = (E_t - E_{t-1})$
$Q =$ producción
$t =$ tiempo

La ecuación anterior se basó en el supuesto de que el nivel deseado de empleo, $E_t^*$, es una función de la producción, del tiempo y del tiempo elevado al cuadrado, y según la hipótesis de que $E_t - E_{t-1} = \delta(E_t^* - E_{t-1})$, donde $\delta$, el coeficiente de ajuste, se encuentra entre 0 y 1.

*a*) Interprete la regresión anterior.

*b*) ¿Cuál es el valor de $\delta$?

*c*) Derive la función de demanda de trabajo de largo plazo a partir de la función de demanda de corto plazo estimada.

*d*) ¿Cómo comprobaría la existencia de la correlación serial en el modelo anterior?

17.15. Griliches estudió la demanda agrícola de tractores con el siguiente modelo:†

$$T_t^* = \alpha X_{1,t-1}^{\beta_1} X_{2,t-1}^{\beta_2}$$

donde $T^* =$ existencias deseadas de tractores
$X_1 =$ precios relativos de los tractores
$X_2 =$ tasa de interés

Con el modelo de ajuste de existencias, obtuvo los siguientes resultados para el periodo 1921-1957:

$$\widehat{\log T_t} = \text{constante} - \underset{(0.051)}{0.218 \log X_{1,t-1}} - \underset{(0.170)}{0.855 \log X_{2,t-1}} + \underset{(0.035)}{0.864 \log T_{t-1}}$$

$$R^2 = 0.987$$

donde las cifras entre paréntesis son los errores estándar estimados.

*a*) ¿Cuál es el coeficiente de ajuste estimado?

*b*) ¿Cuáles son las elasticidades de precio de corto y largo plazos?

*c*) ¿Cuáles son las elasticidades de interés correspondientes?

*d*) ¿Cuáles son las razones para una tasa de ajuste alta o baja en este modelo?

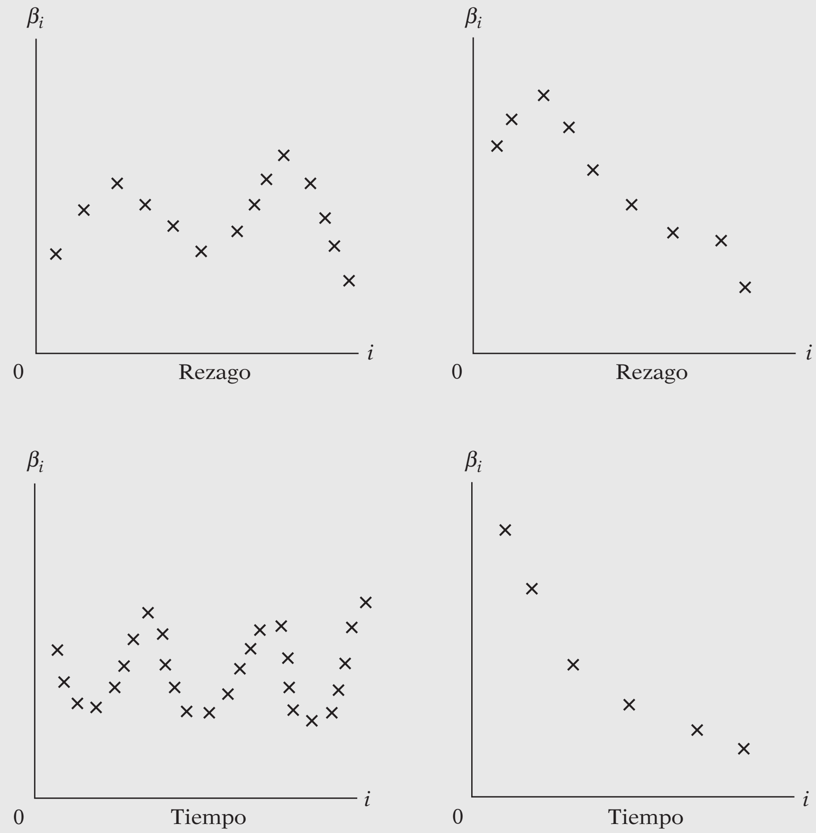
17.16. Siempre que la variable dependiente rezagada aparezca como variable explicativa, la $R^2$ suele ser mucho más alta que cuando no aparece. ¿Cuáles son las razones para este comentario?

---

* F.P.R. Brechling, "The Relationship between Output and Employment in British Manufacturing Industries", *Review of Economic Studies*, vol. 32, julio de 1965.

† Zvi Griliches, "The Demand for a Durable Input: Farm Tractors in the United States, 1921-1957", en Arnold C. Harberger (ed.), *The Demand for Durable Goods*, University of Chicago Press, Chicago, 1960.

**FIGURA 17.10**
Estructuras de rezagos hipotéticas.




17.17. Considere los patrones de rezago de la figura 17.10. ¿A qué grado del polinomio ajustaría las estructuras de rezagos y por qué?

17.18. Considere la ecuación (17.13.4):

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m$$

Obtenemos la varianza de $\hat{\beta}_i$ a partir de las varianzas de $\hat{a}_i$ con la siguiente fórmula:

$$\begin{aligned}\text{var}(\hat{\beta}_i) &= \text{var}(\hat{a}_0 + \hat{a}_1 i + \hat{a}_2 i^2 + \cdots + \hat{a}_m i^m) \\ &= \sum_{j=0}^{m} i^{2j}\,\text{var}(\hat{a}_j) + 2\sum_{j<p} i^{(j+p)}\text{cov}(\hat{a}_j \hat{a}_p)\end{aligned}$$

*a*) Mediante la fórmula anterior, encuentre la varianza de $\hat{\beta}_i$ expresada como

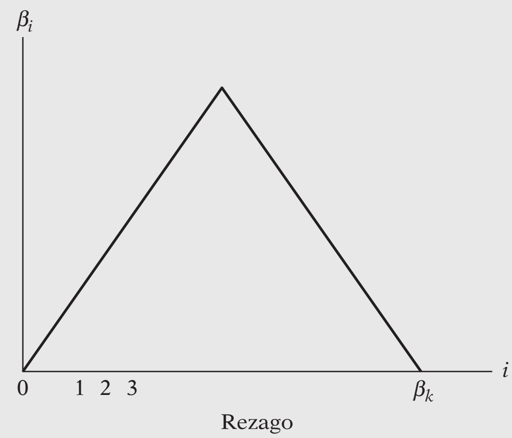
$$\hat{\beta}_i = \hat{a}_0 + \hat{a}_1 i + \hat{a}_2 i^2$$
$$\hat{\beta}_i = \hat{a}_0 + \hat{a}_1 i + \hat{a}_2 i^2 + \hat{a}_3 i^3$$

*b*) Si las varianzas de $\hat{a}_i$ son grandes en relación con ellas mismas, ¿la varianza de $\hat{\beta}_i$ será grande también? ¿Por qué?

17.19. Considere el siguiente modelo de rezagos distribuidos:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

**FIGURA 17.11**
Modelo de rezagos distribuidos en V invertida.




Suponga que $\beta_i$ se expresa adecuadamente mediante el polinomio de segundo grado de la siguiente manera:

$$\beta_i = a_0 + a_1 i + a_2 i^2$$

¿Cómo estimaría las $\beta$ si deseamos imponer la restricción de que $\beta_0 = \beta_4 = 0$?

17.20. *El modelo de rezagos distribuidos en V invertida.* Considere el modelo de rezagos distribuidos finitos de $k$ periodos

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t$$

F. DeLeeuw propuso la estructura para las $\beta$ como la presentada en la figura 17.11, donde las $\beta$ siguen la forma de V invertida. Si suponemos, por simplicidad, que $k$ (la longitud máxima del rezago) es un número par y también que $\beta_0$ y $\beta_k$ son cero, DeLeeuw propone el siguiente esquema para las $\beta$:*

$$\begin{aligned} \beta_i &= i\beta & 0 \le i \le \frac{k}{2} \\ &= (k-i)\beta & \frac{k}{2} \le i < k \end{aligned}$$

¿Cómo utilizaría el esquema de DeLeeuw para estimar los parámetros del modelo anterior de rezagos distribuidos de $k$ periodos?

17.21. Consulte el ejercicio 12.15. Como el valor $d$ que allí aparece es de poca utilidad para detectar la autocorrelación (de primer orden) (¿por qué?), ¿cómo probaría la presencia de autocorrelación en este caso?

## Ejercicios empíricos

17.22. Considere el siguiente modelo:

$$Y_i^* = \alpha + \beta_0 X_t + u_t$$

donde $Y^*$ = el gasto empresarial deseado, o de largo plazo, en nueva planta y equipo, $X_t$ = ventas y $t$ = tiempo. Con el modelo de ajuste de existencias, estime los parámetros

* Véase su artículo "The Demand for Capital Goods by Manufacturers: A Study of Quarterly Time Series", *Econometrica*, vol. 30, núm. 3, julio de 1962, pp. 407-423.

**TABLA 17.10**
**Inversión en planta fija y equipo en el sector manufacturero $Y$ y ventas industriales $X_2$ en miles de millones de dólares, ajustadas por estacionalidad, Estados Unidos, 1970-1991**

Fuente: *Economic Report of the President*, 1993. Los datos sobre $Y$ se tomaron de la tabla B-52, p. 407; los datos sobre $X$ se tomaron de la tabla 8-53, p. 408.

| Año | Gasto en planta, $Y$ | Ventas, $X_2$ | Año | Gasto en planta, $Y$ | Ventas, $X_2$ |
|---|---|---|---|---|---|
| 1970 | 36.99 | 52.805 | 1981 | 128.68 | 168.129 |
| 1971 | 33.60 | 55.906 | 1982 | 123.97 | 163.351 |
| 1972 | 35.42 | 63.027 | 1983 | 117.35 | 172.547 |
| 1973 | 42.35 | 72.931 | 1984 | 139.61 | 190.682 |
| 1974 | 52.48 | 84.790 | 1985 | 152.88 | 194.538 |
| 1975 | 53.66 | 86.589 | 1986 | 137.95 | 194.657 |
| 1976 | 58.53 | 98.797 | 1987 | 141.06 | 206.326 |
| 1977 | 67.48 | 113.201 | 1988 | 163.45 | 223.541 |
| 1978 | 78.13 | 126.905 | 1989 | 183.80 | 232.724 |
| 1979 | 95.13 | 143.936 | 1990 | 192.61 | 239.459 |
| 1980 | 112.60 | 154.391 | 1991 | 182.81 | 235.142 |

de la función de demanda de largo y de corto plazos del gasto en la nueva planta y equipo a partir de la información en la tabla 17.10.

¿Cómo averiguaría si hay correlación serial en la información?

17.23. Con la información del ejercicio 17.22, considere el siguiente modelo:

$$Y_i^* = \beta_0 X_t^{\beta_1} e^{u_t}$$

Con el modelo de ajuste de existencias (¿por qué?), estime las elasticidades de corto y de largo plazos del gasto en nueva planta y equipo respecto de las ventas. Compare los resultados con los obtenidos en el ejercicio 17.22. ¿Qué modelo seleccionaría y por qué? ¿Hay correlación serial en la información? ¿Cómo sabe?

17.24. Con la información del ejercicio 17.22, suponga que

$$Y_t = \alpha + \beta X_t^* + u_t$$

donde $X_t^*$ son las ventas deseadas. Estime los parámetros de este modelo y compare los resultados con los obtenidos en el ejercicio 17.22. ¿Cómo decidiría cuál es el modelo apropiado? Con base en el estadístico $h$, concluiría que hay correlación serial en la información?

17.25. Suponga que alguien lo convence de que la relación entre el gasto de la empresa en una nueva planta y en equipo y las ventas es la siguiente:

$$Y_t^* = \alpha + \beta X_t^* + u_t$$

donde $Y^*$ es el gasto deseado y $X^*$ son las ventas deseadas o esperadas. Con la información del ejercicio 17.22, estime este modelo y comente los resultados.

17.26. Con los datos del ejercicio 17.22, determine si el gasto en planta causa (a la manera de Granger) las ventas o si las ventas causan (a la manera de Granger) el gasto en planta. Utilice hasta seis rezagos y comente los resultados. ¿Qué conclusiones importantes obtiene de este ejercicio?

17.27. Suponga que las ventas en el ejercicio 17.22 tienen un efecto de rezagos distribuidos sobre el gasto por concepto de planta y equipo. Ajuste un modelo de rezagos Almon adecuado para los datos.

17.28. Estime de nuevo la ecuación (17.13.16) imponiendo 1) una restricción sobre el extremo cercano, 2) una restricción sobre el extremo lejano y 3) restricciones en ambos extremos; además, compare los resultados de la ecuación (17.13.16). ¿Qué conclusión general obtiene?

**TABLA 17.11 Inversiones, ventas y tasa de interés, Estados Unidos, 1960-1999**

| Observación | Inversión | Ventas | Interés | Observación | Inversión | Ventas | Interés |
|---|---|---|---|---|---|---|---|
| 1960 | 4.9 | 60 827 | 4.41 | 1980 | 69.6 | 327 233 | 11.94 |
| 1961 | 5.2 | 61 159 | 4.35 | 1981 | 82.4 | 355 822 | 14.17 |
| 1962 | 5.7 | 65 662 | 4.33 | 1982 | 88.9 | 347 625 | 13.79 |
| 1963 | 6.5 | 68 995 | 4.26 | 1983 | 100.8 | 369 286 | 12.04 |
| 1964 | 7.3 | 73 682 | 4.40 | 1984 | 121.7 | 410 124 | 12.71 |
| 1965 | 8.5 | 80 283 | 4.49 | 1985 | 130.8 | 422 583 | 11.37 |
| 1966 | 10.6 | 87 187 | 5.13 | 1986 | 137.6 | 430 419 | 9.02 |
| 1967 | 11.2 | 90 820 | 5.51 | 1987 | 141.9 | 457 735 | 9.38 |
| 1968 | 11.9 | 96 685 | 6.18 | 1988 | 155.9 | 497 157 | 9.71 |
| 1969 | 14.6 | 105 690 | 7.03 | 1989 | 173.0 | 527 039 | 9.26 |
| 1970 | 16.7 | 108 221 | 8.04 | 1990 | 176.1 | 545 909 | 9.32 |
| 1971 | 17.3 | 116 895 | 7.39 | 1991 | 181.4 | 542 815 | 8.77 |
| 1972 | 19.3 | 131 081 | 7.21 | 1992 | 197.5 | 567 176 | 8.14 |
| 1973 | 23.0 | 153 677 | 7.44 | 1993 | 215.0 | 595 628 | 7.22 |
| 1974 | 26.8 | 177 912 | 8.57 | 1994 | 233.7 | 639 163 | 7.96 |
| 1975 | 28.2 | 182 198 | 8.83 | 1995 | 262.0 | 684 982 | 7.59 |
| 1976 | 32.4 | 204 150 | 8.43 | 1996 | 287.3 | 718 113 | 7.37 |
| 1977 | 38.6 | 229 513 | 8.02 | 1997 | 325.2 | 753 445 | 7.26 |
| 1978 | 48.3 | 260 320 | 8.73 | 1998 | 367.4 | 779 413 | 6.53 |
| 1979 | 58.6 | 297 701 | 9.63 | 1999 | 433.0 | 833 079 | 7.04 |

*Notas:* Inversión = inversión fija privada en equipo para el procesamiento de información y software, miles de millones de dólares, ajustada por estacionalidad.
Ventas = ventas totales en industria y comercio, millones de dólares, ajustadas por estacionalidad.
Interés = tasa de los bonos empresariales Aaa de Moody, %.

Fuente: *Economic Report of the President*, 2001, tablas B-18, B-57 y B-73.

17.29. La tabla 17.11 proporciona datos sobre la inversión privada fija en procesamiento de información y equipo ($Y$, en miles de millones de dólares), las ventas totales en industria y comercio ($X_2$ en millones de dólares) y la tasa de interés ($X_3$, tasa de los bonos empresariales Aaa de Moody); los datos sobre $Y$ y $X_2$ están ajustados por estacionalidad.

*a*) Pruebe la causalidad bilateral entre $Y$ y $X_2$, con especial atención a la longitud del rezago.

*b*) Pruebe la causalidad bilateral entre $Y$ y $X_3$, con especial atención también a la longitud del rezago.

*c*) Para permitir un efecto de rezagos distribuidos de las ventas sobre la inversión, suponga que se decide por la técnica de rezagos de Almon. Muestre el modelo estimado después de fijarse con cuidado en la longitud del rezago, así como en el grado del polinomio.

17.30. La tabla 17.12 proporciona los datos sobre índices del salario real por hora ($Y$) y de la productividad por hora ($X_2$), con 1992 = 100 como base de ambos índices en el sector de negocios de la economía estadounidense de 1960 a 1999, así como la tasa de desempleo civil ($X_3$) durante el mismo periodo.

*a*) ¿Cómo decide si la compensación salarial determina la productividad laboral o viceversa?

*b*) Elabore un modelo adecuado para probar la conjetura de *a*), y proporcione las estadísticas usuales.

*c*) ¿Es posible que la tasa de desempleo tenga algún efecto sobre la compensación salarial?, y de ser así, ¿cómo tomaría esto en cuenta? Muestre los análisis estadísticos necesarios.

**TABLA 17.12** **Remuneración, productividad y tasa de desempleo, Estados Unidos, 1960-1999**

| Observación | REMUN | PRODUCT | TasDES | Observación | REMUN | PRODUCT | TasDES |
|---|---|---|---|---|---|---|---|
| 1960 | 60.0 | 48.8 | 5.5 | 1980 | 89.5 | 80.4 | 7.1 |
| 1961 | 61.8 | 50.6 | 6.7 | 1981 | 89.5 | 82.0 | 7.6 |
| 1962 | 63.9 | 52.9 | 5.5 | 1982 | 90.9 | 81.7 | 9.7 |
| 1963 | 65.4 | 55.0 | 5.7 | 1983 | 91.0 | 84.6 | 9.6 |
| 1964 | 67.9 | 57.5 | 5.2 | 1984 | 91.3 | 87.0 | 7.5 |
| 1965 | 69.4 | 59.6 | 4.5 | 1985 | 92.7 | 88.7 | 7.2 |
| 1966 | 71.9 | 62.0 | 3.8 | 1986 | 95.8 | 91.4 | 7.0 |
| 1967 | 73.8 | 63.4 | 3.8 | 1987 | 96.3 | 91.9 | 6.2 |
| 1968 | 76.3 | 65.4 | 3.6 | 1988 | 97.3 | 93.0 | 5.5 |
| 1969 | 77.4 | 65.7 | 3.5 | 1989 | 95.9 | 93.9 | 5.3 |
| 1970 | 78.9 | 67.0 | 4.9 | 1990 | 96.5 | 95.2 | 5.6 |
| 1971 | 80.4 | 69.9 | 5.9 | 1991 | 97.5 | 96.3 | 6.8 |
| 1972 | 82.7 | 72.2 | 5.6 | 1992 | 100.0 | 100.0 | 7.5 |
| 1973 | 84.5 | 74.5 | 4.9 | 1993 | 99.9 | 100.5 | 6.9 |
| 1974 | 83.5 | 73.2 | 5.6 | 1994 | 99.7 | 101.9 | 6.1 |
| 1975 | 84.4 | 75.8 | 8.5 | 1995 | 99.3 | 102.6 | 5.6 |
| 1976 | 86.8 | 78.5 | 7.7 | 1996 | 99.7 | 105.4 | 5.4 |
| 1977 | 87.9 | 79.8 | 7.1 | 1997 | 100.4 | 107.6 | 4.9 |
| 1978 | 89.5 | 80.7 | 6.1 | 1998 | 104.3 | 110.5 | 4.5 |
| 1979 | 89.7 | 80.7 | 5.8 | 1999 | 107.3 | 114.0 | 4.2 |

*Notas:* REMUN = índice de remuneración real por hora (1992 = 100).
PRODUCT = índice de producción por hora (1992 = 100).
TasDES = tasa de desempleo civil, %.

Fuente: *Economic Report of the President*, 2001, tabla B-49, p. 332.

17.31. En una prueba de causalidad de Granger, Christopher Sims explota el hecho de que el futuro no puede causar el presente.* Para decidir si una variable *Y* causa una variable *X,* Sims propone estimar el siguiente par de ecuaciones:

$$Y_t = \alpha_1 + \sum_{i=1}^{i=n} \beta_i X_{t-i} + \sum_{i=1}^{i=m} \gamma_i Y_{t-i} + \sum_{i=1}^{i=p} \lambda_i X_{t+i} + u_{1t} \tag{1}$$

$$X_t = \alpha_2 + \sum_{i=1}^{i=n} \delta_i X_{t-i} + \sum_{i=1}^{i=m} \theta_i Y_{t-i} + \sum_{i=1}^{i=p} \omega_i Y_{t+i} + u_{2t} \tag{2}$$

Estas regresiones incluyen valores rezagados, actuales y futuros, o **adelantados**, de las regresoras; los términos como $X_{t+1}$, $X_{t+2}$, se llaman **términos adelantados**.

Si *Y* es causa según Granger de *X,* debe haber alguna relación entre *Y* y los valores adelantados, o futuros, de *X.* Por tanto, en lugar de probar que $\Sigma\beta_i = 0$, debemos probar que $\Sigma\lambda_i = 0$ en la ecuación (1). Si rechazamos esta hipótesis, la causalidad va de *Y* a *X,* y no de *X* a *Y,* porque el futuro no puede causar el presente. Los mismos comentarios aplican a la ecuación (2).

* C.A. Sims, "Money, Income, and Causality", *American Economic Review*, vol. 62, 1972, pp. 540-552.

**TABLA 17.13**
**Datos macroeconómicos de la economía griega, 1960-1995**

Fuente: H. R. Seddighi, K. A. Lawler y A. V. Katos, *Econometrics: A Practical Approach*, Routledge, Londres, 2000, p. 158.

| Año | CP | IDP | Invbr | PNB | TIL |
|---|---|---|---|---|---|
| 1960 | 107 808 | 117 179 | 29 121 | 145 458 | 8 |
| 1961 | 115 147 | 127 599 | 31 476 | 161 802 | 8 |
| 1962 | 120 050 | 135 007 | 34 128 | 164 674 | 8 |
| 1963 | 126 115 | 142 128 | 35 996 | 181 534 | 8.25 |
| 1964 | 137 192 | 159 649 | 43 445 | 196 586 | 9 |
| 1965 | 147 707 | 172 756 | 49 003 | 214 922 | 9 |
| 1966 | 157 687 | 182 366 | 50 567 | 228 040 | 9 |
| 1967 | 167 528 | 195 611 | 49 770 | 240 791 | 9 |
| 1968 | 179 025 | 204 470 | 60 397 | 257 226 | 8.75 |
| 1969 | 190 089 | 222 638 | 71 653 | 282 168 | 8 |
| 1970 | 206 813 | 246 819 | 70 663 | 304 420 | 8 |
| 1971 | 217 212 | 269 249 | 80 558 | 327 723 | 8 |
| 1972 | 232 312 | 297 266 | 92 977 | 356 886 | 8 |
| 1973 | 250 057 | 335 522 | 100 093 | 383 916 | 9 |
| 1974 | 251 650 | 310 231 | 74 500 | 369 325 | 11.83 |
| 1975 | 266 884 | 327 521 | 74 660 | 390 000 | 11.88 |
| 1976 | 281 066 | 350 427 | 79 750 | 415 491 | 11.5 |
| 1977 | 293 928 | 366 730 | 85 950 | 431 164 | 12 |
| 1978 | 310 640 | 390 189 | 91 100 | 458 675 | 13.46 |
| 1979 | 318 817 | 406 857 | 99 121 | 476 048 | 16.71 |
| 1980 | 319 341 | 401 942 | 92 705 | 485 108 | 21.25 |
| 1981 | 325 851 | 419 669 | 85 750 | 484 259 | 21.33 |
| 1982 | 338 507 | 421 716 | 84 100 | 483 879 | 20.5 |
| 1983 | 339 425 | 417 930 | 83 000 | 481 198 | 20.5 |
| 1984 | 345 194 | 434 696 | 78 300 | 490 881 | 20.5 |
| 1985 | 358 671 | 456 576 | 82 360 | 502 258 | 20.5 |
| 1986 | 361 026 | 439 654 | 77 234 | 507 199 | 20.5 |
| 1987 | 365 473 | 438 454 | 73 315 | 505 713 | 21.82 |
| 1988 | 378 488 | 476 345 | 79 831 | 529 460 | 22.89 |
| 1989 | 394 942 | 492 334 | 87 873 | 546 572 | 23.26 |
| 1990 | 403 194 | 495 939 | 96 139 | 546 982 | 27.62 |
| 1991 | 412 458 | 513 173 | 91 726 | 566 586 | 29.45 |
| 1992 | 420 028 | 502 520 | 93 140 | 568 582 | 28.71 |
| 1993 | 420 585 | 523 066 | 91 292 | 569 724 | 28.56 |
| 1994 | 426 893 | 520 728 | 93 073 | 579 846 | 27.44 |
| 1995 | 433 723 | 518 407 | 98 470 | 588 691 | 23.05 |

*Nota:* Todos los datos nominales se expresan a precios constantes de mercado del año 1970 en millones de dracmas. El ingreso disponible privado se deflaciona por el deflactor de precios al consumidor.

Para realizar la prueba de Sims, estimamos la ecuación (1) sin los términos adelantados (*regresión restringida*) y después estimamos la ecuación (1) con los términos adelantados (*regresión no restringida*). Enseguida realizamos la prueba $F$ como se indica en la ecuación (8.7.1). Si el estadístico $F$ es significativo (por ejemplo, en el nivel de 5%), concluimos que es $Y$ la que causa, según Granger, $X$. Los mismos comentarios aplican a la ecuación (2).

¿Qué prueba elegimos, la de Granger o la de Sims? Podemos aplicar ambas.* El factor favorable de la prueba de Granger es que necesita menos grados de libertad porque no usa

* La elección entre las pruebas de causalidad de Granger y Sims no es clara. Para un análisis más detallado de estas pruebas, véase G. Chamberlain, "The General Equivalence of Granger and Sims Causality", *Econometrica*, vol. 50, 1982, pp. 569-582.

los términos adelantados. Si la muestra no es lo bastante grande, habrá que tener cuidado al usar la prueba de Sims.

Consulte los datos del ejercicio 12.34. Para efectos pedagógicos, aplique la prueba de causalidad de Sims para determinar si las ventas causan el gasto en planta o viceversa. Use los datos de los últimos cuatro años como términos adelantados del análisis.

17.32 La tabla 17.13 presenta algunos datos macroeconómicos de la economía griega de 1960 a 1995.

Considere la siguiente función de consumo:

$$\ln \mathrm{CP}_t^* = \beta_1 + \beta_2 \ln \mathrm{IDP}_t + \beta_3 \mathrm{TIL}_t + u_t$$

donde $\mathrm{CP}_t^*$ = gasto de consumo privado deseado real en el tiempo *t;* $\mathrm{IDP}_t$ = ingreso disponible privado en el tiempo *t*; $\mathrm{TIL}_t$ = tasa de interés de largo plazo en el tiempo *t*; y ln representa el logaritmo natural.

*a*) Con base en los datos de la tabla 17.13, estime la función de consumo anterior e indique claramente cómo midió el gasto de consumo privado deseado real.

*b*) ¿Qué problemas econométricos se presentaron al estimar la función de consumo anterior? ¿Cómo los resolvió? Proporcione una explicación detallada.

17.33 Con los datos de la tabla 17.13 formule un modelo adecuado para explicar el comportamiento de la inversión real bruta en la economía griega de 1960 a 1995. Busque en cualquier libro de macroeconomía el modelo acelerador de inversión.

# Apéndice 17A

## 17A.1 Prueba de Sargan para la validez de los instrumentos

Suponga que utilizamos una o varias variables instrumentales para sustituir variables explicativas correlacionadas con el término de error. ¿Qué validez tiene(n) la(s) variable(s) instrumental(es)?, es decir, ¿cómo sabe que los instrumentos elegidos son independientes del término de error? Sargan desarrolló un estadístico, llamado SARG, para probar la validez de los instrumentos utilizados en variables instrumentales (VI).* El proceso SARG consiste en los siguientes pasos:†

1. Divida en dos grupos las variables incluidas en una ecuación de regresión: las independientes del término de error (por ejemplo, $X_1, X_2, \ldots, X_p$), y las no independientes del término de error (es decir, $Z_1, Z_2, \ldots, Z_q$).
2. Sean $W_1, W_2, \ldots, W_s$ los instrumentos elegidos para las variables $Z$ en 1, donde $s > q$.
3. Calcule la regresión original, sustituyendo las $Z$ por las $W$; es decir, estime la regresión original mediante la(s) VI y obtenga los residuos, por ejemplo $\hat{u}$.
4. Haga la regresión de $\hat{u}$ sobre una constante, todas las variables $X$ y sobre las variables $W$, pero excluya todas las variables $Z$. Obtenga $R^2$ de esta regresión.
5. Ahora calcule el estadístico SARG, definido como:

$$\mathrm{SARG} = (n-k)R^2 \sim \chi^2_{s-q} \qquad \textbf{(17A.1.1)}$$

* J.D. Sargan, "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", en P.E. Hart, G. Mills y J.K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Butterworths, Londres, 1964.

† El siguiente análisis se basa en H.R. Seddighi, K.A. Lawler y A.V. Katos, *Econometrics: A Practical Approach*, Routledge, Nueva York, 2000, pp. 155-156.

donde $n =$ número de observaciones y $k =$ número de coeficientes en la ecuación de regresión original. Según la hipótesis nula de que los instrumentos son exógenos, Sargan demostró que la prueba SARG asintóticamente sigue la distribución $\chi^2$ con $(s - q)$ grados de libertad, donde $s$ es el número de instrumentos (es decir, las variables en $W$) y $q$ es el número de regresoras en la ecuación original. Si el valor calculado de ji cuadrada en una aplicación es significativo estadísticamente, rechazamos la validez de los instrumentos. Si no es significativo estadísticamente, aceptamos como válido el instrumento elegido. Debe subrayarse que $s > q$, es decir, el número de instrumentos debe ser mayor que $q$. Si no es así (es decir, $s \leq q$), la prueba SARG no es válida.

6. La hipótesis nula es que todos los instrumentos ($W$) son válidos. Si la ji cuadrada calculada excede el valor ji cuadrada crítico, rechazamos la hipótesis nula, lo cual significa que al menos un instrumento está correlacionado con el término de error y, por tanto, las estimaciones de la(s) VI con base en los instrumentos elegidos no son válidas.

# Modelos de ecuaciones simultáneas y econometría de series de tiempo

# Parte 4

Una mirada informal al trabajo empírico publicado en administración de empresas y en economía revelará que muchas relaciones económicas son de tipo uniecuacional. Por esta razón dedicamos las tres primeras partes de este libro a modelos de regresión uniecuacionales. En tales modelos, una variable (la variable dependiente $Y$) se expresa como función lineal de una o más variables (las variables explicativas, las $X$). En tales modelos, un supuesto implícito es que la relación causa-efecto, de existir, entre $Y$ y $X$ es unidireccional: las variables explicativas son la *causa* y la variable dependiente es el *efecto*.

Sin embargo, hay situaciones en las cuales existe una influencia bidireccional entre las variables económicas; es decir, una variable económica afecta otra(s) variable(s) económica(s) y, a su vez, se ve afectada por ella(s). Así, en la regresión del dinero $M$ sobre la tasa de interés $r$, la metodología uniecuacional supone implícitamente que la tasa de interés es fija (por ejemplo, por el Sistema de la Reserva Federal) y trata de encontrar la respuesta del dinero demandado a cambios en el nivel de la tasa de interés. Pero, ¿qué sucede si la tasa de interés depende de la demanda de dinero? En este caso, el análisis de regresión condicional en este libro, hasta el momento, quizá no sea apropiado porque ahora $M$ depende de $r$ y $r$ depende de $M$. Por tanto, es preciso considerar dos ecuaciones, una que relaciona $M$ con $r$ y otra que relaciona $r$ con $M$. Y esto conduce a la consideración de los modelos de ecuaciones simultáneas, modelos en los cuales hay más de una ecuación de regresión, una por cada variable interdependiente.

En la **parte 4** presentamos una introducción muy elemental y casi heurística al complejo tema de **modelos de ecuaciones simultáneas**, y dejamos los detalles para las referencias.

En el capítulo 18 ofrecemos diversos ejemplos de modelos de ecuaciones simultáneas y veremos por qué el método de mínimos cuadrados ordinarios, considerado antes, es inaplicable para estimar los parámetros de cada ecuación del modelo.

En el capítulo 19 consideraremos el **problema de identificación**. Si en un sistema de ecuaciones simultáneas con dos o más ecuaciones no es posible obtener valores numéricos de cada parámetro en cada ecuación porque las ecuaciones son *observacionalmente no distinguibles,* o se parecen mucho entre sí, tenemos el problema de identificación. Así, en la regresión de la cantidad $Q$ sobre el precio $P$, ¿es la ecuación resultante una función de demanda o una función

de oferta, pues $Q$ y $P$ forman parte de las dos funciones? Por consiguiente, si sólo tenemos información sobre $Q$ y $P$, y no hay otra información, será difícil, si no imposible, identificar la regresión como función de demanda o función de oferta. Es indispensable resolver el problema de identificación antes de proceder a la estimación, pues no saber lo que se estima hace que la estimación misma carezca de sentido. En el capítulo 19 se ofrecen diversos métodos para resolver el problema de la identificación.

En el capítulo 20 consideraremos diversos métodos de estimación diseñados específicamente para estimar los modelos de ecuaciones simultáneas, así como sus cualidades y limitaciones.

# Capítulo 18

# Modelos de ecuaciones simultáneas

En este capítulo y en los dos siguientes analizaremos los modelos de ecuaciones simultáneas. En particular, observaremos sus características especiales, su estimación y algunos problemas estadísticos relacionados con ellos.

## 18.1 Naturaleza de los modelos de ecuaciones simultáneas

En las **partes 1** a **3** de este texto tratamos exclusivamente con modelos uniecuacionales, es decir, modelos en los cuales había una sola variable dependiente $Y$ y una o más variables explicativas, las $X$. En tales modelos nos centramos en la estimación y/o la predicción del valor medio de $Y$ condicional a los valores fijos de las variables $X$. Por consiguiente, la relación causa-efecto en esos modelos iba de las $X$ a $Y$.

Pero en muchas situaciones la relación causa-efecto en un sentido, o unidireccional, no es relevante. Esto sucede cuando $Y$ está determinada por las $X$ y algunas $X$, a su vez, están determinadas por $Y$. En otras palabras, hay una relación en dos sentidos, o simultánea, entre $Y$ y (algunas) $X$, que hace dudar del valor de la distinción entre variables *dependientes* y *explicativas*. Es mejor reunir un conjunto de variables que se determinen simultáneamente mediante el conjunto restante de variables: justo lo que se hace en los modelos de ecuaciones simultáneas. En tales modelos hay más de una ecuación: una para cada una de las **variables** *mutuamente*, o *conjuntamente*, dependientes o **endógenas**.[1] Y, a diferencia de los modelos uniecuacionales, en los modelos de ecuaciones simultáneas no es posible estimar los parámetros de una ecuación aisladamente sin tener en cuenta la información proporcionada por las demás ecuaciones en el sistema.

¿Qué sucede si los parámetros de cada ecuación se estiman, por ejemplo, mediante el método de MCO, sin considerar las demás ecuaciones del sistema? Recuerde que un supuesto crucial del método de MCO es que las variables explicativas $X$ son no estocásticas o, si lo son (aleatorias), están distribuidas independientemente del término de perturbación estocástico. Si ninguna de estas condiciones se cumple, entonces, como veremos más adelante, los estimadores de mínimos cuadrados no sólo son sesgados, sino también inconsistentes; es decir, a medida que el

[1] En el contexto de los modelos de ecuaciones simultáneas, las variables conjuntamente dependientes se denominan **variables endógenas,** y las variables realmente no estocásticas o que pueden considerarse como tales, **variables exógenas** o **predeterminadas**. (Veremos más al respecto en el capítulo 19.)

tamaño de la muestra aumenta indefinidamente, los estimadores no convergen hacia sus verdaderos valores (poblacionales). Así, en el siguiente sistema hipotético de ecuaciones,[2]

$$Y_{1i} = \beta_{10} + \beta_{12}Y_{2i} + \gamma_{11}X_{1i} + u_{1i} \tag{18.1.1}$$

$$Y_{2i} = \beta_{20} + \beta_{21}Y_{1i} + \gamma_{21}X_{1i} + u_{2i} \tag{18.1.2}$$

donde $Y_1$ y $Y_2$ son variables mutuamente dependientes, o endógenas, $X_1$, una variable exógena, y $u_1$ y $u_2$, los términos de perturbación estocástica, ambas variables $Y_1$ y $Y_2$ son estocásticas. Por consiguiente, a menos que se demuestre que la variable explicativa estocástica $Y_2$ en (18.1.1) está distribuida independientemente de $u_1$ y que la variable explicativa estocástica $Y_1$ en (18.1.2) está distribuida independientemente de $u_2$, la aplicación de MCO clásicos a estas ecuaciones generará estimaciones inconsistentes.

En lo que resta de este capítulo daremos ejemplos de modelos de ecuaciones simultáneas y mostraremos el sesgo en la aplicación directa del método de mínimos cuadrados a tales modelos. Después de analizar el denominado problema de identificación en el capítulo 19, en el 20 estudiaremos algunos métodos especiales para manejar los modelos de ecuaciones simultáneas.

## 18.2 Ejemplos de modelos de ecuaciones simultáneas

**EJEMPLO 18.1**
*Modelo de demanda y oferta*

Como es bien sabido, el precio *P* de un bien y la cantidad vendida *Q* están determinados por el intercepto de las curvas de demanda y oferta para ese bien. Así, si suponemos, por simplicidad, que las curvas de demanda y oferta son lineales y sumamos los términos de perturbación estocásticos $u_1$ y $u_2$, las funciones empíricas de demanda y oferta se escriben como

*Función de demanda:* $$Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t} \qquad \alpha_1 < 0 \tag{18.2.1}$$

*Función de oferta:* $$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t} \qquad \beta_1 > 0 \tag{18.2.2}$$

*Condición de equilibrio:* $$Q_t^d = Q_t^s$$

donde $Q^d$ = cantidad demandada
$Q^s$ = cantidad ofrecida
$t$ = tiempo

y las $\alpha$ y $\beta$ son los parámetros. *A priori*, se espera que $\alpha_1$ sea negativa (curva de demanda con pendiente hacia abajo) y que $\beta_1$ sea positiva (curva de oferta con pendiente hacia arriba).

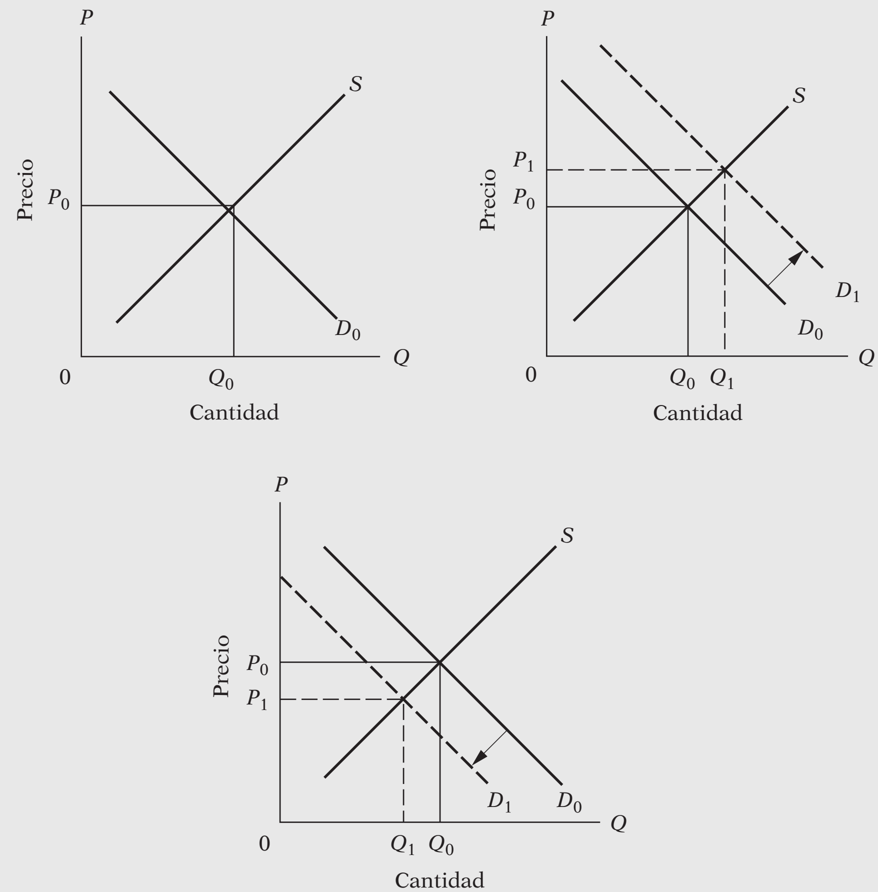
Ahora bien, no es muy difícil ver que *P* y *Q* son variables conjuntamente dependientes. Si, por ejemplo, $u_{1t}$ en (18.2.1) se modifica debido a cambios en otras variables que afectan a $Q_t^d$ (como ingreso, riqueza y gustos), la curva de demanda se desplaza hacia arriba si $u_{1t}$ es positiva y hacia abajo si $u_{1t}$ es negativa. Estos desplazamientos se muestran en la figura 18.1.

Como muestra la figura, un desplazamiento en la curva de demanda cambia a *P* y a *Q*. En forma similar, un cambio en $u_{2t}$ (huelgas, clima, restricciones sobre las importaciones o las exportaciones, etc.) desplazará la curva de oferta, para afectar de nuevo a *P* y a *Q*. Debido a esta dependencia simultánea entre *Q* y *P*, $u_{1t}$ y $P_t$ en (18.2.1) y $U_{2t}$ y $P_t$ en (18.2.2) no pueden ser independientes. Por consiguiente, una regresión de *Q* sobre *P* como en (18.2.1) violaría un supuesto importante del modelo clásico de regresión lineal, a saber, el de no correlación entre la(s) variable(s) explicativa(s) y el término de perturbación.

[2] Esta notación económica, aunque se explica por sí misma, se generalizará a más de dos ecuaciones en el capítulo 19.

**EJEMPLO 18.1** (*continuación*)

**FIGURA 18.1** Interdependencia entre precio y cantidad.




**EJEMPLO 18.2**
*Modelo keynesiano de determinación del ingreso*

Considere el modelo keynesiano simple de determinación del ingreso:

*Función consumo:* $C_t = \beta_0 + \beta_1 Y_t + u_t \qquad 0 < \beta_1 < 1$ **(18.2.3)**

*Identidad del ingreso:* $Y_t = C_t + I_t (= S_t)$ **(18.2.4)**

donde
$C =$ gasto de consumo
$Y =$ ingreso
$I =$ inversión (se supone exógena)
$S =$ ahorro
$t =$ tiempo
$u =$ término de perturbación estocástico
$\beta_0$ y $\beta_1$ = parámetros

El parámetro $\beta_1$ se conoce como *propensión marginal a consumir* (PMC) (la cantidad de gasto de consumo extra resultante de un dólar extra de ingreso). Conforme a la teoría económica, se espera que $\beta_1$ se encuentre entre 0 y 1. La ecuación (18.2.3) es la función consumo (estocástica); y (18.2.4) es la identidad del ingreso nacional, que significa que el ingreso total es igual al gasto

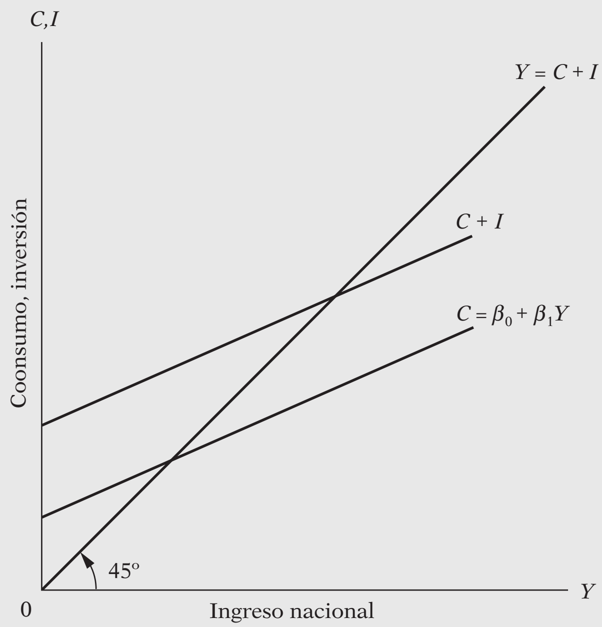
(*continúa*)

**EJEMPLO 18.2**
(*continuación*)

de consumo total más el gasto de inversión total, en el entendido de que el gasto de inversión total es igual al ahorro total. En la figura 18.2 se muestra la gráfica.

De la función consumo postulada y de la figura 18.2, es claro que $C$ y $Y$ son interdependientes y que no se espera que $Y_t$ en (18.2.3) sea independiente del término de perturbación, porque cuando $u_t$ se desplaza (debido a diversos factores dentro del término de error), la función consumo también se desplaza, la cual a su vez afecta a $Y_t$. Por consiguiente, una vez más, el método clásico de mínimos cuadrados no es aplicable a (18.2.3). De aplicarse, los estimadores obtenidos de dicho método serán inconsistentes, como veremos más adelante.

**FIGURA 18.2**
Modelo keynesiano de determinación del ingreso.




**EJEMPLO 18.3**
*Modelos de salario-precio*

Considere el siguiente modelo tipo Phillips de determinación de salarios monetarios y precios:

$$\dot{W}_t = \alpha_0 + \alpha_1 \text{UN}_t + \alpha_2 \dot{P}_t + u_{1t} \qquad \textbf{(18.2.5)}$$

$$\dot{P}_t = \beta_0 + \beta_1 \dot{W}_t + \beta_2 \dot{R}_t + \beta_3 \dot{M}_t + u_{2t} \qquad \textbf{(18.2.6)}$$

donde $\dot{W}$ = tasa de cambio de los salarios monetarios
UN = tasa de desempleo, %
$\dot{P}$ = tasa de cambio de los precios
$\dot{R}$ = tasa de cambio del costo de capital
$\dot{M}$ = tasa de cambio del precio de las materias primas importadas
$t$ = tiempo
$u_1, u_2$ = perturbaciones estocásticas

Como la variable precio $\dot{P}$ entra en la ecuación de salarios y la variable salarios $W$ entra en la ecuación de precios, las dos variables son conjuntamente dependientes. Por consiguiente, se espera que estas variables explicativas estocásticas estén correlacionadas con las perturbaciones estocásticas pertinentes, por lo que, una vez más, el método clásico de MCO no es inaplicable para estimar individualmente los parámetros de las dos ecuaciones.

**EJEMPLO 18.4**
*El modelo IS de macroeconomía*

El conocido modelo IS, o de equilibrio en el mercado de bienes, de la macroeconomía[3] en su forma no estocástica se expresa como

| | | | |
|---|---|---|---|
| *Función consumo:* | $C_t = \beta_0 + \beta_1 Y_{dt}$ | $0 < \beta_1 < 1$ | **(18.2.7)** |
| *Función de impuestos:* | $T_t = \alpha_0 + \alpha_1 Y_t$ | $0 < \alpha_1 < 1$ | **(18.2.8)** |
| *Función de inversión:* | $I_t = \gamma_0 + \gamma_1 r_t$ | | **(18.2.9)** |
| *Definición:* | $Y_{dt} = Y_t - T_t$ | | **(18.2.10)** |
| *Gasto del gobierno:* | $G_t = \bar{G}$ | | **(18.2.11)** |
| *Identidad del ingreso nacional:* | $Y_t = C_t + I_t + G_t$ | | **(18.2.12)** |

donde $Y =$ ingreso nacional
$C =$ gasto de consumo
$I =$ inversión neta planeada o deseada
$\bar{G} =$ nivel dado de gasto del gobierno
$T =$ impuestos
$Y_d =$ ingreso disponible
$r =$ tasa de interés

Si sustituye (18.2.10) y (18.2.8) en (18.2.7) y también la ecuación resultante por $C$, así como las ecuaciones (18.2.9) y (18.2.11) en (18.2.12), debe obtener la ecuación IS:

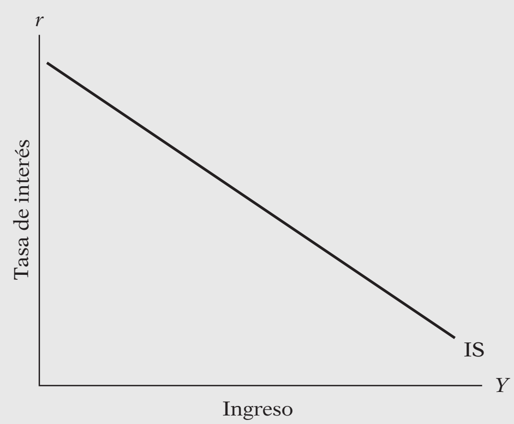
$$Y_t = \pi_0 + \pi_1 r_t \tag{18.2.13}$$

donde

$$\pi_0 = \frac{\beta_0 - \alpha_0\beta_1 + \gamma_0 + \bar{G}}{1 - \beta_1(1 - \alpha_1)}$$
$$\pi_1 = \frac{1}{1 - \beta_1(1 - \alpha_1)} \tag{18.2.14}$$

La ecuación (18.2.13) es la ecuación de IS, o de equilibrio en el mercado de bienes, es decir, da las combinaciones de tasa de interés y de nivel de ingreso de modo que el mercado de bienes se despeja o está en equilibrio. Geométricamente, la curva IS se muestra en la figura 18.3.

**FIGURA 18.3**
Curva IS.




(*continúa*)

[3] "El esquema de equilibrio en el mercado de bienes, o esquema IS, muestra combinaciones de tasas de interés y de niveles de producto tales que el gasto planeado iguala al ingreso." Véase Rudiger Dornbusch y Stanley Fischer, *Macroeconomics*, 3a. ed., McGraw-Hill, Nueva York, 1984, p. 102. Observe que, por simplicidad, supusimos que no existe el sector de comercio exterior.

**EJEMPLO 18.4**
(*continuación*)

¿Qué sucedería si estimamos, por ejemplo, la función consumo (18.2.7) en forma aislada? ¿Obtendríamos estimaciones insesgadas y/o consistentes de $\beta_0$ y $\beta_1$? Tal resultado no es probable, pues el consumo depende del ingreso disponible, el cual depende del ingreso nacional $Y$, que a su vez depende de $r$ y $\bar{G}$, como también de otros parámetros que entran en $\pi_0$. Por consiguiente, a menos que consideremos todas estas influencias, es probable que una simple regresión de $C$ sobre $Y_d$ produzca estimaciones sesgadas y/o inconsistentes de $\beta_0$ y $\beta_1$.

**EJEMPLO 18.5**
*Modelo LM*

La otra mitad del famoso paradigma IS-LM es el LM, o relación de equilibrio en el mercado monetario, que da las combinaciones de tasa de interés y nivel de ingreso de modo que se despeje el mercado monetario, es decir, que la demanda de dinero sea igual a su oferta. Algebraicamente, el modelo, en la forma no estocástica, se expresa como:

*Función de demanda de dineron:* $$M_t^d = a + bY_t - cr_t \qquad \textbf{(18.2.15)}$$

*Función de oferta de dinero:* $$M_t^s = \bar{M} \qquad \textbf{(18.2.16)}$$

*Condición de equilibrio:* $$M_t^d = M_t^s \qquad \textbf{(18.2.17)}$$

donde $Y =$ ingreso, $r =$ tasa de interés y $\bar{M} =$ nivel supuesto de oferta monetaria, por ejemplo, el determinado por el Banco de la Reserva Federal.

Al igualar las funciones de demanda y oferta de dinero, y simplificar, obtenemos la e*cuación LM*:

$$Y_t = \lambda_0 + \lambda_1 \bar{M} + \lambda_2 r_t \qquad \textbf{(18.2.18)}$$

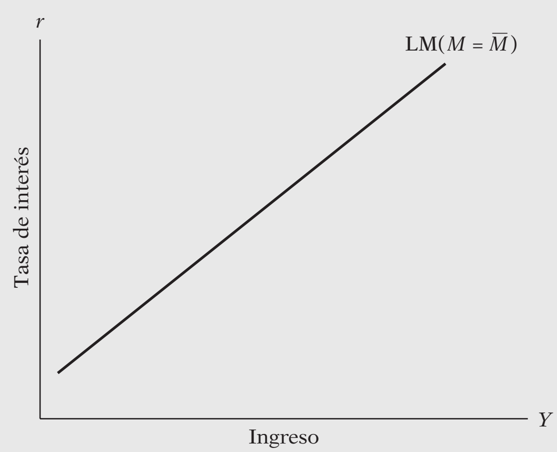
donde

$$\begin{aligned} \lambda_0 &= -a/b \\ \lambda_1 &= 1/b \\ \lambda_2 &= c/b \end{aligned} \qquad \textbf{(18.2.19)}$$

Para un $M = \bar{M}$ dado, la curva LM que representa la relación (18.2.18) es como se muestra en la figura 18.4.

Las curvas IS y LM muestran que un ordenamiento completo de tasas de interés es consistente con el equilibrio en el mercado de bienes, y un ordenamiento completo de tasas de interés es compatible con el equilibrio en el mercado monetario. Desde luego, sólo una tasa de interés y un nivel de ingreso serán consistentes al mismo tiempo con los dos equilibrios. Para obtenerlos, sólo se necesita igualar (18.2.13) y (18.2.18). En el ejercicio 18.4 se le pide mostrar el nivel de la tasa de interés y del ingreso simultáneamente compatible con el equilibrio en los mercados de bienes y de dinero.

**FIGURA 18.4**
Curva LM.



**EJEMPLO 18.6**
*Modelos econométricos*

Los modelos de ecuaciones simultáneas son muy comunes para elaborar modelos econométricos a cargo de diversos econometristas. Un pionero en este campo fue el profesor Lawrence Klein, de la Wharton School, Universidad de Pensilvania. Su modelo inicial, conocido como **modelo 1 de Klein**, es el siguiente:

$$\begin{aligned}
&\textit{Función consumo:} && C_t = \beta_0 + \beta_1 P_t + \beta_2 (W + W')_t + \beta_3 P_{t-1} + u_{1t} \\
&\textit{Función de inversión:} && I_t = \beta_4 + \beta_5 P_t + \beta_6 P_{t-1} + \beta_7 K_{t-1} + u_{2t} \\
&\textit{Demanda de trabajo:} && W_t = \beta_8 + \beta_9 (Y + T - W')_t \\
& && \quad + \beta_{10} (Y + T - W')_{t-1} + \beta_{11} t + u_{3t} \\
&\textit{Identidad:} && Y_t + T_t = C_t + I_t + G_t \\
&\textit{Identidad:} && Y_t = W'_t + W_t + P_t \\
&\textit{Identidad:} && K_t = K_{t-1} + I_t
\end{aligned} \tag{18.2.20}$$

donde
$C$ = gasto de consumo
$I$ = gasto de inversión
$G$ = gasto del gobierno
$P$ = utilidades
$W$ = nómina del sector privado
$W'$ = nómina del gobierno
$K$ = existencias de capital
$T$ = impuestos
$Y$ = ingreso después de impuestos
$t$ = tiempo
$u_1$, $u_2$ y $u_3$ = perturbaciones estocásticas[4]

En el modelo anterior, las variables $C$, $I$, $W$, $Y$, $P$ y $K$ se consideran variables conjuntamente dependientes o endógenas, y las variables $P_{t-1}$, $K_{t-1}$ y $Y_{t-1}$, predeterminadas.[5] En total hay seis ecuaciones (con las tres identidades) para estudiar la interdependencia de las seis variables endógenas.

En el capítulo 20 veremos la forma de estimar tales modelos econométricos. Por el momento, observe que, debido a la interdependencia entre las variables endógenas, en general no son independientes de los términos de perturbación estocásticos, lo cual, por consiguiente, hace que no sea adecuada la aplicación del método de MCO a una ecuación individual en el sistema. Como se ve en la sección 18.3, los estimadores así obtenidos son inconsistentes; no convergen a sus verdaderos valores poblacionales aunque el tamaño de la muestra sea muy grande.

## 18.3 Sesgo en las ecuaciones simultáneas: inconsistencia de los estimadores de MCO

Como ya planteamos, el método de mínimos cuadrados no aplica para estimar una sola ecuación enlazada a un sistema de ecuaciones simultáneas si una o más de las variables explicativas están correlacionadas con el término de perturbación en esa ecuación, porque los estimadores así obtenidos son inconsistentes. Para mostrar esto, considere de nuevo el modelo keynesiano simple de

[4] L.R. Klein, *Economic Fluctuations in the United States, 1921-1941*, John Wiley & Sons, Nueva York, 1950.

[5] El constructor de modelos debe especificar las variables endógenas y predeterminadas del modelo. $K_{t-1}$ y $Y_{t-1}$ son predeterminadas porque, en el tiempo $t$, sus valores son conocidos. (Veremos más sobre esto en el capítulo 19.)

determinación del ingreso del ejemplo 18.2. Suponga que deseamos estimar los parámetros de la función consumo (18.2.3). Si suponemos que $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_{t+j}) = 0$ (para $j \neq 0$ ) y $\text{cov}(I_t, u_t) = 0$, que son los supuestos del MCRL, demostramos primero que $Y_t$ y $u_t$ en (18.2.3) están correlacionados y luego probamos que $\hat{\beta}_1$ es un estimador inconsistente de $\beta_1$.

Para probar que $Y_t$ y $u_t$ están correlacionados, procedemos de la siguiente manera. Sustituya (18.2.3) en (18.2.4) para obtener

$$Y_t = \beta_0 + \beta_1 Y_t + u_t + I_t$$

es decir,

$$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t \tag{18.3.1}$$

Ahora

$$E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t \tag{18.3.2}$$

donde aprovechamos que $E(u_t) = 0$ y que, como $I_t$ es exógeno o predeterminado (porque su valor se fijó con anterioridad), tiene como valor esperado $I_t$.

Por consiguiente, al restar (18.3.2) de (18.3.1), resulta

$$Y_t - E(Y_t) = \frac{u_t}{1-\beta_1} \tag{18.3.3}$$

Además,

$$u_t - E(u_t) = u_t \qquad \text{(¿Por qué?)} \tag{18.3.4}$$

de donde

$$\begin{aligned} \text{cov}(Y_t, u_t) &= E[Y_t - E(Y_t)][u_t - E(u_t)] \\ &= \frac{E\left(u_t^2\right)}{1-\beta_1} \qquad \text{con (18.3.3) y (18.3.4)} \\ &= \frac{\sigma^2}{1-\beta_1} \end{aligned} \tag{18.3.5}$$

Como $\sigma^2$ se supuso positivo (¿por qué?), la covarianza entre $Y$ y $u$ dada en (18.3.5) tiende a ser diferente de cero.[6] Como resultado se espera que $Y_t$ y $u_t$ en (18.2.3) estén correlacionadas, lo cual viola el supuesto del modelo clásico de regresión lineal respecto de que las perturbaciones son independientes o por lo menos no están correlacionadas con las variables explicativas. Como ya mencionamos, los estimadores de MCO en esta situación son inconsistentes.

Para mostrar que el estimador de MCO $\hat{\beta}_1$ es un estimador inconsistente de $\beta_1$ debido a la correlación entre $Y_t$ y $u_t$, procedemos de la siguiente manera:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum(C_t - \bar{C})(Y_t - \bar{Y})}{\sum(Y_t - \bar{Y})^2} \\ &= \frac{\sum c_t y_t}{\sum y_t^2} \\ &= \frac{\sum C_t y_t}{\sum y_t^2} \end{aligned} \tag{18.3.6}$$

[6] Será mayor que cero siempre que $\beta_1$, la PMC, se encuentre entre 0 y 1; y será negativa si $\beta_1$ es mayor que la unidad. Desde luego, un valor de PMC mayor que la unidad no tendría mucho sentido económico. En realidad, se espera que la covarianza entre $Y_t$ y $u_t$ sea positiva.

donde las letras minúsculas, como es usual, indican desviaciones de la media (muestras). Al sustituir por $C_t$ de (18.2.3), obtenemos

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum(\beta_0 + \beta_1 Y_t + u_t)y_t}{\sum y_t^2} \\ &= \beta_1 + \frac{\sum y_t u_t}{\sum y_t^2}\end{aligned} \tag{18.3.7}$$

donde, en el último paso, aprovechamos que $\sum y_t = 0$ y $(\sum Y_t y_t / \sum y_t^2) = 1$ (¿por qué?).

Si tomamos el valor esperado de (18.3.7) en ambos lados, obtenemos

$$E(\hat{\beta}_1) = \beta_1 + E\left[\frac{\sum y_t u_t}{\sum y_t^2}\right] \tag{18.3.8}$$

Por desgracia, no podemos evaluar $E(\sum y_t u_t / \sum y_t^2)$ porque el operador de valor esperado es un operador lineal. [*Nota:* $E(A/B) \neq E(A)/E(B)$.] Pero, por intuición, debe quedar claro que, a menos que el término $(\sum y_t u_t / \sum y_t^2)$ sea cero, $\hat{\beta}_1$ es un estimador sesgado de $\beta_1$. Pero, ¿no demostramos en (18.3.5) que la covarianza entre $Y$ y $u$ es diferente de cero y que, por consiguiente, $\hat{\beta}_1$ no estaría sesgado? La respuesta es no del todo, pues cov($Y_t$, $u_t$), un concepto poblacional, no equivale exactamente a $\sum y_t u_t$, que es una medición muestral, aunque, a medida que el tamaño de la muestra aumenta indefinidamente, el último tenderá hacia el primero. Pero si el tamaño de la muestra aumenta indefinidamente, entonces podemos recurrirse al concepto de estimador consistente y averiguar qué sucede con $\hat{\beta}_1$ a medida que $n$, el tamaño de la muestra, aumenta indefinidamente. En resumen, cuando no podemos evaluar explícitamente el valor esperado de un estimador, como ocurrió en (18.3.8), podemos centrar la atención hacia su comportamiento en una muestra grande.

Ahora bien, se dice que un estimador es consistente si el **límite de su probabilidad**,[7] o **plím** para abreviar, es igual a su verdadero valor (poblacional). Por consiguiente, para demostrar que $\hat{\beta}_1$ de (18.3.7) es inconsistente, debemos demostrar que su plim no es igual al verdadero $\beta_1$. Al aplicar las reglas de límite de probabilidad a (18.3.7), obtenemos[8]

$$\begin{aligned}\text{plím}\,(\hat{\beta}_1) &= \text{plím}\,(\beta_1) + \text{plím}\left(\frac{\sum y_t u_t}{\sum y_t^2}\right) \\ &= \text{plím}\,(\beta_1) + \text{plím}\left(\frac{\sum y_t u_t / n}{\sum y_t^2 / n}\right) \\ &= \beta_1 + \frac{\text{plím}\left(\sum y_t u_t / n\right)}{\text{plím}\left(\sum y_t^2 / n\right)}\end{aligned} \tag{18.3.9}$$

donde, en el segundo paso, dividimos $\sum y_t u_t$ y $\sum y_t^2$ entre el número total de observaciones en la muestra, $n$, de manera que las cantidades en los paréntesis son ahora la covarianza muestral entre $Y$ y $u$, y la varianza muestral de $Y$, respectivamente.

En palabras, (18.3.9) establece que el límite de probabilidad de $\hat{\beta}_1$ es igual al verdadero $\beta_1$ más la razón del plim de la covarianza muestral entre $Y$ y $u$ respecto del plim de la varianza muestral de $Y$. Ahora, a medida que el tamaño $n$ de la muestra aumenta indefinidamente, se esperaría que la covarianza muestral entre $Y$ y $u$ se aproxime a la verdadera covarianza poblacional $E[Y_t - E(Y_t)][u_t - E(u_t)]$, la cual, de (18.3.5), es igual a $[\sigma^2/(1 - \beta_1)]$. En forma similar, a me-

[7] En el **apéndice A** definimos el límite de probabilidad.

[8] Como afirmamos en el **apéndice A**, el plim de una constante (por ejemplo, $\beta_1$) es la constante misma, y el plim de $(A/B) = \text{plim}(A)/\text{plim}(B)$. Observe, sin embargo, que $E(A/B) \neq E(A)/E(B)$.

dida que $n$ tiende a infinito, la varianza muestral de $Y$ se aproxima a su varianza poblacional, es decir, $\sigma_Y^2$. Por consiguiente, la ecuación (18.3.9) puede escribirse como

$$\begin{aligned} \text{plím}\,(\hat{\beta}_1) &= \beta_1 + \frac{\sigma^2/(1-\beta_1)}{\sigma_Y^2} \\ &= \beta_1 + \frac{1}{1-\beta_1}\left(\frac{\sigma^2}{\sigma_Y^2}\right) \end{aligned} \tag{18.3.10}$$

Como $0 < \beta_1 < 1$ y $\sigma^2$, y $\sigma_Y^2$ son positivas, es obvio, de la ecuación (18.3.10), que plím ($\hat{\beta}_1$) será siempre mayor que $\beta_1$; es decir, $\hat{\beta}_1$ sobreestimará al verdadero $\hat{\beta}_1$.[9] En otras palabras, $\hat{\beta}_1$ es un estimador sesgado, y no importa lo grande del tamaño de la muestra, el sesgo no desaparecerá.

## 18.4 Sesgo de las ecuaciones simultáneas: ejemplo numérico

Para demostrar algunos puntos planteados en la sección anterior, considere de nuevo el modelo keynesiano simple de determinación del ingreso dado en el ejemplo 18.2 y efectúe el siguiente estudio de **Monte Carlo**.[10] Suponga que los valores de la inversión I son como se muestran en la columna (3) de la tabla 18.1, y además que

$$\begin{aligned} E(u_t) &= 0 \\ E(u_t u_{t+j}) &= 0 \qquad (j \neq 0) \\ \text{var}\,(u_t) &= \sigma^2 = 0.04 \\ \text{cov}\,(u_t, I_t) &= 0 \end{aligned}$$

Los $u_t$ así generados se muestran en la columna (4).

Para la función consumo (18.2.3) suponga que se conocen los valores de los verdaderos parámetros y son $\beta_0 = 2$ y $\beta_1 = 0.8$.

De los valores supuestos de $\beta_0$ y $\beta_1$ y de los valores generados de $u_t$ podemos generar los valores del ingreso $Y_t$ de (18.3.1), los cuales se muestran en la columna (1) de la tabla 18.1. Una vez conocidos los $Y_t$ y al conocer $\beta_0$, $\beta_1$ y $u_t$, podemos generar fácilmente los valores de consumo $C_t$ de (18.2.3). Los $C$ así generados están en la columna 2.

Como se conocen los verdaderos $\beta_0$ y $\beta_1$ y los errores muestrales son exactamente los mismos que los "verdaderos" (debido a la forma en que se diseñó el estudio Monte Carlo), si utilizamos la información de la tabla 18.1 para hacer la regresión de $C_t$ sobre $Y_t$, debemos obtener $\beta_0 = 2$ y $\beta_1 = 0.8$, si los MCO fueran insesgados. Pero, de (18.3.7), sabemos que éste no será el caso si la regresora $Y_t$ y la perturbación $u_t$ están correlacionadas. Ahora no es muy difícil verificar, de la información disponible, que la covarianza (muestral) entre $Y_t$ y $u_t$ es $\sum y_t u_t, = 3.8$ y que $\sum y_t^2 = 184$. Entonces, como indica (18.3.7), debemos tener

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \frac{\sum y_t u_t}{\sum y_t^2} \\ &= 0.8 + \frac{3.8}{184} \\ &= 0.82065 \end{aligned} \tag{18.4.1}$$

Es decir, $\hat{\beta}_1$ está sesgado hacia arriba por 0.02065.

[9] En general, sin embargo, la dirección del sesgo depende de la estructura del modelo particular y de los verdaderos valores de los coeficientes de la regresión.

[10] Tomado de Kenneth J. White, Nancy G. Horsman y Justin B. Wyatt, *SHAZAM: Computer Handbook for Econometric for Use with Basic Econometrics*, McGraw-Hill, Nueva York, pp. 131-134.

**TABLA 18.1**

| $Y_t$ (1) | $C_t$ (2) | $I_t$ (3) | $u_t$ (4) |
|---|---|---|---|
| 18.15697 | 16.15697 | 2.0 | −0.3686055 |
| 19.59980 | 17.59980 | 2.0 | −0.8004084E-01 |
| 21.93468 | 19.73468 | 2.2 | 0.1869357 |
| 21.55145 | 19.35145 | 2.2 | 0.1102906 |
| 21.88427 | 19.48427 | 2.4 | −0.2314535E-01 |
| 22.42648 | 20.02648 | 2.4 | 0.8529544E-01 |
| 25.40940 | 22.80940 | 2.6 | 0.4818807 |
| 22.69523 | 20.09523 | 2.6 | −0.6095481E-01 |
| 24.36465 | 21.56465 | 2.8 | 0.7292983E-01 |
| 24.39334 | 21.59334 | 2.8 | 0.7866819E-01 |
| 24.09215 | 21.09215 | 3.0 | −0.1815703 |
| 24.87450 | 21.87450 | 3.0 | −0.2509900E-01 |
| 25.31580 | 22.11580 | 3.2 | −0.1368398 |
| 26.30465 | 23.10465 | 3.2 | 0.6092946E-01 |
| 25.78235 | 22.38235 | 3.4 | −0.2435298 |
| 26.08018 | 22.68018 | 3.4 | −0.1839638 |
| 27.24440 | 23.64440 | 3.6 | −0.1511200 |
| 28.00963 | 24.40963 | 3.6 | 0.1926739E-02 |
| 30.89301 | 27.09301 | 3.8 | 0.3786015 |
| 28.98706 | 25.18706 | 3.8 | −0.2588852E-02 |

Fuente: Kenneth J. White, Nancy G. Horsman y Justin B. Wyatt, *SHAZAM Computer Handbook for Econometrics for Use with Damodar Gujarati: Basic Econometrics*, septiembre, 1985, p. 132.

Ahora efectuemos la regresión de $C_t$ sobre $Y_t$ con la información de la tabla 18.1; los resultados de la regresión son

$$\begin{aligned} \hat{C}_t &= 1.4940 \quad + \quad 0.82065Y_t \\ \text{ee} &= (0.35413) \quad (0.01434) \\ t &= (4.2188) \quad (57.209) \qquad R^2 = 0.9945 \end{aligned} \tag{18.4.2}$$

Como se esperaba, la $\beta_1$ estimada es precisamente la predicha por (18.4.1). A propósito, observe que la $\beta_0$ estimada también está sesgada.

En general, el valor del sesgo en $\hat{\beta}_1$ depende de $\beta_1$, $\sigma^2$ y var($Y$), y, en particular, del grado de la covarianza entre $Y$ y $u$.[11] Como afirman Kenneth White *et al.*, "en esto consiste el sesgo de las ecuaciones simultáneas. En contraste con los modelos uniecuacionales, ya no podemos seguir suponiendo que las variables del lado derecho de la ecuación no están correlacionadas con el término de error".[12] Tenga en cuenta que este sesgo permanece aun en muestras grandes.

En vista de las consecuencias potencialmente graves de la aplicación del MCO a los modelos de ecuaciones simultáneas, ¿existe una prueba de simultaneidad que indique si en un momento dado se tiene un problema de simultaneidad? Una versión de la **prueba de especificación de Hausman** sirve para este propósito, y la analizaremos en el capítulo 19.

[11] Véase la ecuación (18.3.5).

[12] *Op. cit.*, pp. 133-134.

## Resumen y conclusiones

1. En contraste con los modelos uniecuacionales, los de ecuaciones simultáneas contienen más de una variable dependiente, o **endógena**, lo cual requiere un número de ecuaciones igual al número de variables endógenas.
2. Una característica única de los modelos de ecuaciones simultáneas es que la variable endógena (es decir, la variable regresada) en una ecuación puede aparecer como variable explicativa (es decir, como regresora) en otra ecuación del sistema.
3. Como consecuencia, tal **variable explicativa endógena** se convierte en estocástica y suele estar correlacionada con el término de perturbación de la ecuación en la cual aparece como variable explicativa.
4. En esta situación no es aplicable el método clásico de MCO porque los estimadores así obtenidos no son consistentes, es decir, no convergen hacia sus verdaderos valores poblacionales sin importar qué tan grande sea la muestra.
5. El ejemplo de Monte Carlo presentado en el texto muestra la naturaleza del sesgo contenido en la aplicación de MCO para estimar los parámetros de una ecuación de regresión, en la cual la regresora está correlacionada con el término de perturbación, que es el caso habitual en los modelos de ecuaciones simultáneas.
6. Como los modelos de ecuaciones simultáneas son frecuentes, sobre todo en los modelos econométricos, diversos autores han desarrollado técnicas alternas de estimación. Las analizaremos en el capítulo 20, después del **problema de identificación** en el capítulo 19, tema que lógicamente es previo a la estimación.

## EJERCICIOS

### Preguntas

18.1. Elabore un modelo de ecuaciones simultáneas para la oferta y la demanda de odontólogos en Estados Unidos. Especifique las variables endógenas y exógenas en el modelo.

18.2. Elabore un modelo simple de la demanda y la oferta de dinero en Estados Unidos, y compárelo con los de K. Brunner y A. H. Meltzer,* y R. Tiegen.†

18.3. *a*) Para el modelo de demanda y oferta del ejemplo 18.1, obtenga la expresión para el límite de probabilidad de $\hat{\alpha}_1$.

*b*) ¿En qué condiciones este límite de probabilidad será igual al verdadero $\alpha_1$?

18.4. Para el modelo IS-LM analizado en el texto, encuentre los niveles de tasa de interés y de ingreso simultáneamente compatibles con el equilibrio del mercado de bienes y de dinero.

18.5. Para estudiar la relación entre la inflación y el rendimiento de las acciones comunes, Bruno Oudet‡ utilizó el siguiente modelo:

$$R_{bt} = \alpha_1 + \alpha_2 R_{st} + \alpha_3 R_{bt-1} + \alpha_4 L_t + \alpha_5 Y_t + \alpha_6 \text{NIS}_t + \alpha_7 I_t + u_{1t}$$
$$R_{st} = \beta_1 + \beta_2 R_{bt} + \beta_3 R_{bt-1} + \beta_4 L_t + \beta_5 Y_t + \beta_6 \text{NIS}_t + \beta_7 E_t + u_{2t}$$

* "Some Further Evidence on Supply and Demand Functions for Money", *Journal of Finance*, vol. 19, mayo de 1964, pp. 240-283.

† "Demand and Supply Functions for Money in the United States", *Econometrica*, vol. 32, núm. 4, octubre de 1964, pp. 476-509.

‡ Bruno A. Oudet, "The Variation of the Return on Stocks in Periods of Inflation", *Journal of Financial and Quantitative Analysis*, vol. 8, núm. 2, marzo de 1973, pp. 247-258.

donde $L =$ base monetaria real *per cápita*
$Y =$ ingreso real *per cápita*
$I =$ tasa de inflación esperada
$\text{NIS} =$ variable de una nueva emisión
$E =$ rendimientos esperados de acciones a fin de periodo, representados por razones de precios de acciones rezagadas
$R_{bt} =$ rendimiento de los bonos
$R_{st} =$ rendimiento de las acciones comunes

*a*) Presente una justificación teórica para este modelo y vea si su razonamiento coincide con el de Oudet.

*b*) ¿Cuáles son las variables endógenas del modelo? ¿Y las exógenas?

*c*) ¿Cómo consideraría las $R_{bt}$ rezagadas endógenas o exógenas?

18.6. En su artículo, "Un modelo de distribución de productos de uso personal de marca en Jamaica",* John U. Farley y Harold J. Levitt desarrollaron el siguiente modelo (los productos de uso personal fueron crema de afeitar, crema para la piel, pañales desechables y crema dental):

$$Y_{1i} = \alpha_1 + \beta_1 Y_{2i} + \beta_2 Y_{3i} + \beta_3 Y_{4i} + u_{1i}$$
$$Y_{2i} = \alpha_2 + \beta_4 Y_{1i} + \beta_5 Y_{5i} + \gamma_1 X_{1i} + \gamma_2 X_{2i} + u_{2i}$$
$$Y_{3i} = \alpha_3 + \beta_6 Y_{2i} + \gamma_3 X_{3i} + u_{3i}$$
$$Y_{4i} = \alpha_4 + \beta_7 Y_{2i} + \gamma_4 X_{4i} + u_{4i}$$
$$Y_{5i} = \alpha_5 + \beta_8 Y_{2i} + \beta_9 Y_{3i} + \beta_{10} Y_{4i} + u_{5i}$$

donde $Y_1 =$ porcentaje de tiendas que tienen existencias del producto
$Y_2 =$ unidades vendidas por mes
$Y_3 =$ índice de contacto directo con el importador y con el fabricante del producto
$Y_4 =$ índice de actividad de las ventas al mayoreo en el área
$Y_5 =$ índice de penetración de marca del producto en existencia (por ejemplo, número promedio de marcas de un mismo producto almacenado que mantienen las tiendas que ofrecen el producto en venta)
$X_1 =$ población objetivo para el producto
$X_2 =$ ingreso *per cápita* en la población donde se sitúa el área
$X_3 =$ distancia del centro de gravedad poblacional a Kingston
$X_4 =$ distancia del centro poblacional al pueblo mayorista más cercano

*a*) ¿Puede identificar las variables endógenas y exógenas en el modelo anterior?

*b*) ¿Puede estimar una o más ecuaciones en el modelo mediante el método de mínimos cuadrados? ¿Por qué?

18.7. Para estudiar la relación entre el gasto en publicidad y las ventas de cigarrillos, Frank Bass utilizó el siguiente modelo:†

$$Y_{1t} = \alpha_1 + \beta_1 Y_{3t} + \beta_2 Y_{4t} + \gamma_1 X_{1t} + \gamma_2 X_{2t} + u_{1t}$$
$$Y_{2t} = \alpha_2 + \beta_3 Y_{3t} + \beta_4 Y_{4t} + \gamma_3 X_{1t} + \gamma_4 X_{2t} + u_{2t}$$
$$Y_{3t} = \alpha_3 + \beta_5 Y_{1t} + \beta_6 Y_{2t} + u_{3t}$$
$$Y_{4t} = \alpha_4 + \beta_7 Y_{1t} + \beta_8 Y_{2t} + u_{4t}$$

* *Journal of Marketing Research*, noviembre de 1968, pp. 362-368.

† "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes", *Journal of Marketing Research*, vol. 6, agosto de 1969, pp. 291-300.

donde $Y_1$ = logaritmo de las ventas de cigarrillos con filtro (número de cigarrillos) dividido entre la población mayor de 20 años

$Y_2$ = logaritmo de ventas de cigarrillos sin filtro (número de cigarrillos) dividido entre la población mayor de 20 años

$Y_3$ = logaritmo del valor de la publicidad de cigarrillos con filtro en dólares dividido entre la población mayor de 20 años, dividido a su vez entre el índice de precios de la publicidad

$Y_4$ = logaritmo del valor de la publicidad de cigarrillos sin filtro en dólares dividido entre la población mayor de 20 años, dividido a su vez entre el índice de precios de la publicidad

$X_1$ = logaritmo del ingreso personal disponible dividido entre la población mayor de 20 años, dividido a su vez entre el índice de precios al consumidor

$X_2$ = logaritmo del precio por paquete de cigarrillos sin filtro dividido entre el índice de precios al consumidor

*a*) En el modelo anterior, las $Y$ son endógenas y las $X$ son exógenas. ¿Por qué supone el autor que $X_2$ es exógena?

*b*) Si $X_2$ se considera una variable endógena, ¿cómo modificaría el modelo anterior?

18.8. G. Menges desarrolló el siguiente modelo econométrico para la economía de Alemania Occidental:*

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 I_t + u_{1t}$$

$$I_t = \beta_3 + \beta_4 Y_t + \beta_5 Q_t + u_{2t}$$

$$C_t = \beta_6 + \beta_7 Y_t + \beta_8 C_{t-1} + \beta_9 P_t + u_{3t}$$

$$Q_t = \beta_{10} + \beta_{11} Q_{t-1} + \beta_{12} R_t + u_{4t}$$

donde $Y$ = ingreso nacional

$I$ = formación neta de capital

$C$ = consumo personal

$Q$ = utilidades

$P$ = índice del costo de vida

$R$ = productividad industrial

$t$ = tiempo

$u$ = perturbaciones estocásticas

*a*) ¿Qué variables consideraría endógenas y cuáles exógenas?

*b*) ¿Hay alguna ecuación en el sistema que pueda estimarse mediante el método de mínimos cuadrados uniecuacional?

*c*) ¿Cuál es la razón para incluir la variable $P$ en la función consumo?

18.9. L. E. Gallaway y P. E. Smith elaboraron un modelo simple para la economía de Estados Unidos, que es el siguiente:†

$$Y_t = C_t + I_t + G_t$$

$$C_t = \beta_1 + \beta_2 \text{YD}_{t-1} + \beta_3 M_t + u_{1t}$$

$$I_t = \beta_4 + \beta_5 (Y_{t-1} - Y_{t-2}) + \beta_6 Z_{t-1} + u_{2t}$$

$$G_t = \beta_7 + \beta_8 G_{t-1} + u_{3t}$$

* G. Menges, "Ein Ökonometriches Modell der Bundesrepublik Deutschland (Vier Strukturgleichungen)", *I.F.O. Studien*, vol. 5, 1959, pp. 1-22.

† "A Quarterly Econometric Model of the United States", *Journal of American Statistical Association*, vol. 56, 1961, pp. 379-383.

donde $Y$ = producto nacional bruto
$C$ = gasto de consumo personal
$I$ = inversión privada doméstica bruta
$G$ = gasto del gobierno más inversión extranjera neta
YD = ingreso disponible, o después de impuestos
$M$ = oferta monetaria al principio del trimestre
$Z$ = ingreso patrimonial antes de impuestos
$t$ = tiempo
$u_1$, $u_2$ y $u_3$ = perturbaciones estocásticas

Todas las variables están medidas en forma de primeras diferencias.

Con base en la información trimestral de 1948 a 1957, los autores aplicaron el método de mínimos cuadrados a cada ecuación individualmente y obtuvieron los siguientes resultados:

$$\hat{C}_t = 0.09 + 0.43\text{YD}_{t-1} + 0.23M_t \qquad R^2 = 0.23$$

$$\hat{I}_t = 0.08 + 0.43(Y_{t-1} - Y_{t-2}) + 0.48Z_t \qquad R^2 = 0.40$$

$$\hat{G}_t = 0.13 + 0.67G_{t-1} \qquad R^2 = 0.42$$

*a*) ¿Cómo justifica el uso del método de mínimos cuadrados uniecuacional en este caso?

*b*) ¿Por qué los valores $R^2$ son relativamente bajos?

## Ejercicios empíricos

18.10. En la tabla 18.2 se da la siguiente información sobre $Y$ (producto interno bruto), $C$ (gasto de consumo personal) e $I$ (inversión privada doméstica bruta), en miles de millones de dólares de 1996, en Estados Unidos, de 1970 a 2006. Suponga que $C$ está relacionada linealmente con $Y$ como en el modelo keynesiano simple de determinación del ingreso del ejemplo 18.2. Obtenga estimaciones por MCO de los parámetros de la función consumo. Guarde los resultados para una revisión posterior, con los métodos desarrollados en el capítulo 20.

18.11. Con la información del ejercicio 18.10, efectúe la regresión de la inversión doméstica bruta $I$ sobre el PIB y guarde los resultados para examinarlos de nuevo en un capítulo posterior.

18.12. Considere la identidad macroeconómica

$$C + I = Y \qquad (= \text{PIB})$$

Igual que antes, suponga que

$$C_t = \beta_0 + \beta_1 Y_t + u_t$$

y, según el **modelo acelerador** de macroeconomía, sea

$$I_t = \alpha_0 + \alpha_1 (Y_t - Y_{t-1}) + v_t$$

donde $u$ y $v$ son los términos de error. Con la información del ejercicio 18.10, estime el modelo acelerador y guarde los resultados para un estudio posterior.

18.13. *Oferta y demanda de gasolina.* La tabla 18.3, que se encuentra en el sitio web del libro de texto, presenta datos sobre algunas variables que determinaron la demanda y oferta de gasolina en Estados Unidos de enero de 1978 a agosto de 2002.* Las variables son

* Estos datos se tomaron del sitio web de Stephen J. Schmidt, *Econometrics*, McGraw-Hill, Nueva York, 2005, www.mhhe.com/economics.

**TABLA 18.2** **Gasto de consumo personal, inversión privada doméstica bruta y PIB, Estados Unidos, 1970-2006 (miles de millones de dólares de 1996)**

| Observación | *C* | *I* | *Y* | Observación | *C* | *I* | *Y* |
|---|---|---|---|---|---|---|---|
| 1970 | 2 451.9 | 427.1 | 3 771.9 | 1989 | 4 675.0 | 926.2 | 6 981.4 |
| 1971 | 2 545.5 | 475.7 | 3 898.6 | 1990 | 4 770.3 | 895.1 | 7 112.5 |
| 1972 | 2 701.3 | 532.1 | 4 105.0 | 1991 | 4 778.4 | 822.2 | 7 100.5 |
| 1973 | 2 833.8 | 594.4 | 4 341.5 | 1992 | 4 934.8 | 889.0 | 7 336.6 |
| 1974 | 2 812.3 | 550.6 | 4 319.6 | 1993 | 5 099.8 | 968.3 | 7 532.7 |
| 1975 | 2 876.9 | 453.1 | 4 311.2 | 1994 | 5 290.7 | 1 099.6 | 7 835.5 |
| 1976 | 3 035.5 | 544.7 | 4 540.9 | 1995 | 5 433.5 | 1 134.0 | 8 031.7 |
| 1977 | 3 164.1 | 627.0 | 4 750.5 | 1996 | 5 619.4 | 1 234.3 | 8 328.9 |
| 1978 | 3 303.1 | 702.6 | 5 015.0 | 1997 | 5 831.8 | 1 387.7 | 8 703.5 |
| 1979 | 3 383.4 | 725.0 | 5 173.4 | 1998 | 6 125.8 | 1 524.1 | 9 066.9 |
| 1980 | 3 374.1 | 645.3 | 5 161.7 | 1999 | 6 438.6 | 1 642.6 | 9 470.3 |
| 1981 | 3 422.2 | 704.9 | 5 291.7 | 2000 | 6 739.4 | 1 735.5 | 9 817.0 |
| 1982 | 3 470.3 | 606.0 | 5 189.3 | 2001 | 6 910.4 | 1 598.4 | 9 890.7 |
| 1983 | 3 668.6 | 662.5 | 5 423.8 | 2002 | 7 099.3 | 1 557.1 | 10 048.8 |
| 1984 | 3 863.3 | 857.7 | 5 813.6 | 2003 | 7 295.3 | 1 613.1 | 10 301.0 |
| 1985 | 4 064.0 | 849.7 | 6 053.7 | 2004 | 7 561.4 | 1 770.2 | 10 675.8 |
| 1986 | 4 228.9 | 843.9 | 6 263.6 | 2005 | 7 803.6 | 1 869.3 | 11 003.4 |
| 1987 | 4 369.8 | 870.0 | 6 475.1 | 2006 | 8 044.1 | 1 919.5 | 11 319.4 |
| 1988 | 4 546.9 | 890.5 | 6 742.7 | | | | |

*Notas:* $C$ = gasto de consumo personal
$I$ = inversión privada doméstica bruta
$Y$ = producto interno bruto (PIB)

Fuente: *Economic Report of the President*, 2001, tabla B-2, p. 276.

pricegas (centavos por galón); quantgas (miles de barriles al día, sin plomo); persincome (ingreso personal, miles de millones de dólares); y ventas de automóviles (millones de automóviles al año).

*a*) Elabore un modelo apropiado de oferta y demanda para el consumo de gasolina.

*b*) ¿Qué variables del modelo de *a*) son endógenas y cuáles exógenas?

*c*) Si estima las funciones de demanda y oferta que obtuvo por MCO, ¿los resultados serán confiables? ¿Por qué?

*d*) Guarde las estimaciones de MCO de las funciones de demanda y oferta para otro ejercicio después de estudiar el capítulo 20.

18.14. La tabla 18.4, que se encuentra en el sitio web del libro de texto, presenta datos macroeconómicos sobre diversas variables de la economía estadounidense para los periodos trimestrales de I-1951 a IV-2000.* Las variables son las siguientes: *Year* = fecha; *Qtr* = trimestre; *Realgdp* = PIB real (miles de millones de dólares); *Realcons* = gasto de consumo real; *Realinvs* = inversión real del sector privado; *Realgovt* = gasto gubernamental real; *Realdpi* = ingreso personal disponible real; *CPI_U* = índice de precios al consumidor; *M*1 = existencia de dinero nominal; *Tbilrate* = promedio trimestral de la tasa de interés de los pagarés de la Tesorería a 90 días, a fin de mes; *Pop* = población, millones, interpolación de cifras a fin de año con tasa de crecimiento constante por trimestre; *Infl* = tasa de inflación (falta la primera observación); y *Realint* = tasa de interés real ex-post = *Tbilrate-Infl* (falta la primera observación).

Con estos datos, formule un modelo macroeconómico sencillo de la economía estadounidense. En el capítulo 20 se le pedirá estimar este modelo.

* Estos datos provienen del Departamento de Comercio, Oficina de Análisis Económico de Estados Unidos, y de www.economagic.com, y se reproducen de William H. Greene, *Econometric Analysis*, 6a. ed., 2008, tabla F5.1, p. 1083.

Capítulo 19

# El problema de la identificación

En este capítulo se considera la naturaleza y el significado del problema de la identificación, cuya esencia es la siguiente: recuerde el modelo de demanda y oferta presentado en la sección 18.2. Suponga que se tiene información de series de tiempo sobre *Q* y *P* solamente y que no hay datos adicionales (tales como el ingreso del consumidor, el precio prevaleciente en el periodo anterior y las condiciones del clima). El problema de la identificación consiste en buscar una respuesta a la siguiente pregunta: dada solamente la información sobre *P* y *Q*, ¿cómo se sabe si se está estimando la función de demanda o la función de oferta? O, dicho de otra manera, si se piensa que se está ajustando una función de demanda, ¿cómo se garantiza que, en realidad, se está estimando dicha función y no otra?

Un momento de reflexión revelará que es necesario responder a la pregunta anterior antes de proceder a estimar los parámetros de la función de demanda. Para resolver el problema de la identificación, primero se introducen algunas notaciones y definiciones, y luego se ilustra dicho problema con diversos ejemplos. En seguida se establecen las reglas que pueden utilizarse para averiguar si una ecuación en un modelo de ecuaciones simultáneas está identificada, es decir, si en realidad se trata de la relación que se está estimando, bien sea la función de demanda, de oferta u otra cualquiera.

## 19.1 Notación y definiciones

Para facilitar la exposición, se introducen las siguientes notaciones y definiciones.

El modelo general de *M* ecuaciones con *M* variables endógenas o conjuntamente dependientes puede escribirse como la ecuación (19.1.1):

$$\begin{aligned}
Y_{1t} &= \qquad\qquad \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \cdots + \beta_{1M}Y_{Mt} \\
&\qquad + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \cdots + \gamma_{1K}X_{Kt} + u_{1t} \\
Y_{2t} &= \beta_{21}Y_{1t} \qquad\quad + \beta_{23}Y_{3t} + \cdots + \beta_{2M}Y_{Mt} \\
&\qquad + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \cdots + \gamma_{2K}X_{Kt} + u_{2t} \\
Y_{3t} &= \beta_{31}Y_{1t} + \beta_{32}Y_{2t} \qquad + \cdots + \beta_{3M}Y_{Mt} \\
&\qquad + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \cdots + \gamma_{3K}X_{Kt} + u_{3t} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_{MT} &= \beta_{M1}Y_{1t} + \beta_{M2}Y_{2t} + \cdots + \beta_{M,M-1}Y_{M-1,t} \\
&\qquad + \gamma_{M1}X_{1t} + \gamma_{M2}X_{2t} + \cdots + \gamma_{MK}X_{Kt} + u_{Mt}
\end{aligned} \tag{19.1.1}$$

donde $Y_1, Y_2, \ldots, Y_M = M$ variables endógenas o conjuntamente dependientes
$X_1, X_2, \ldots, X_K = K$ variables predeterminadas (una de estas variables $X$ puede tomar un valor unitario para dar cabida al término del intercepto en cada ecuación)
$u_1, u_2, \ldots, u_M = M$ perturbaciones estocásticas
$t = 1, 2, \ldots, T =$ número total de observaciones
$\beta =$ coeficientes de las variables endógenas
$\gamma =$ coeficientes de las variables predeterminadas

Como se puede observar, no es preciso que todas y cada una de las variables aparezcan en cada ecuación. En efecto, como se muestra en la sección 19.2, para que una ecuación esté identificada no es indispensable que la totalidad de las variables aparezca en cada ecuación.

En la ecuación (19.1.1) se observa que las variables que forman parte del modelo de ecuaciones simultáneas son de dos tipos: **endógenas**, es decir, aquellas determinadas (cuyos valores están) dentro del modelo; y **predeterminadas**, es decir, aquellas determinadas (cuyos valores están) fuera del modelo. Las variables endógenas se consideran estocásticas, en tanto que las predeterminadas se consideran como no estocásticas.

Las variables predeterminadas están divididas en dos categorías: **exógenas**, tanto actuales como rezagadas, y **endógenas rezagadas**. Así, $X_{1t}$ *es una variable exógena actual (del tiempo presente), mientras que* $X_{1(t-1)}$ es una variable exógena rezagada, con un rezago de un intervalo de tiempo. $Y_{(t-1)}$ es una variable endógena rezagada con rezago de un intervalo de tiempo, pero, puesto que el valor de $Y_{1(t-1)}$ es conocido en el periodo actual $t$, ésta es considerada como no estocástica y, por tanto, es una variable predeterminada.[1] En resumen, las variables exógenas actuales y rezagadas y las endógenas rezagadas se consideran predeterminadas; sus valores no están determinados por el modelo en el periodo de tiempo actual.

Corresponde al diseñador del modelo especificar cuáles variables son endógenas y cuáles son predeterminadas. Aunque las variables (no económicas), tales como la temperatura y la lluvia, son claramente exógenas o predeterminadas, el diseñador de modelos debe tener gran precaución al clasificar las variables económicas como endógenas o predeterminadas, debiendo defender la clasificación con argumentos teóricos *a priori*. No obstante, más adelante en el capítulo se proporciona una prueba estadística de exogeneidad.

Las ecuaciones que aparecen en (19.1.1) se conocen como ecuaciones **estructurales** o **de comportamiento**, porque muestran la estructura (de un modelo económico) de una economía o del comportamiento de un agente económico (por ejemplo, un consumidor o un productor). Las $\beta$ y las $\gamma$ se conocen como **parámetros** o **coeficientes estructurales**.

A partir de las ecuaciones estructurales se pueden resolver para las $M$ variables endógenas, derivar **las ecuaciones en forma reducida** y los correspondientes **coeficientes en forma reducida**. **Una ecuación en forma reducida es aquella que expresa únicamente una variable endógena en términos de las variables predeterminadas y las perturbaciones estocásticas**. A modo de ilustración, puede considerarse el modelo keynesiano de determinación del ingreso presentado en el capítulo 18:

*Función de consumo*: $$C_t = \beta_0 + \beta_1 Y_t + u_t \qquad 0 < \beta_1 < 1 \qquad (18.2.3)$$

*Identidad del ingreso*: $$Y_t = C_t + I_t \qquad (18.2.4)$$

En este modelo, $C$ (consumo) y $Y$ (ingreso) son las variables endógenas e $I$ (gasto de inversión) es considerada como una variable exógena. Ambas ecuaciones son estructurales, siendo la ecuación (18.2.4) una identidad. Como es usual, se supone que la PMC $\beta_1$ se encuentra entre 0 y 1.

Si la ecuación (18.2.3) es sustituida en la ecuación (18.2.4), mediante un reordenamiento algebraico simple se obtiene:

$$Y_t = \Pi_0 + \Pi_1 I_t + w_t \qquad (19.1.2)$$

[1] Se supone implícitamente aquí que las perturbaciones estocásticas, las $u$, no están serialmente correlacionadas. De no ser así, $Y_{t-1}$ estaría correlacionada con el término de perturbación del periodo actual $u_t$ y, por tanto, no se podría tratar como predeterminada.

en donde

$$\Pi_0 = \frac{\beta_0}{1 - \beta_1}$$

$$\Pi_1 = \frac{1}{1 - \beta_1} \qquad \textbf{(19.1.3)}$$

$$w_t = \frac{u_t}{1 - \beta_1}$$

La ecuación (19.1.2) es una **ecuación en forma reducida**; expresa la variable endógena Y solamente como función de la variable exógena $I$ (o predeterminada) y del término de perturbación estocástica $u$. $\Pi_0$ y $\Pi_1$ son los correspondientes **coeficientes en forma reducida**. Observe que estos coeficientes son combinaciones no lineales del (los) coeficiente(s) estructural(es).

Al sustituir el valor de $Y$ de la ecuación (19.1.2) en $C$ de la ecuación (18.2.3), se obtiene otra ecuación en forma reducida:

$$C_t = \Pi_2 + \Pi_3 I_t + w_t \qquad \textbf{(19.1.4)}$$

en donde

$$\Pi_2 = \frac{\beta_0}{1 - \beta_1} \qquad \Pi_3 = \frac{\beta_1}{1 - \beta_1}$$

$$w_t = \frac{u_t}{1 - \beta_1} \qquad \textbf{(19.1.5)}$$

Los coeficientes en la forma reducida, tales como $\Pi_1$ y $\Pi_3$, también se conocen como **multiplicadores de impacto** o **de corto plazo**, porque miden el impacto inmediato sobre la variable endógena de un cambio unitario del valor de la variable exógena.[2] Si en el modelo keynesiano anterior se incrementa el gasto de inversión en 1 dólar, por ejemplo, y se supone que la PMC es de 0.8, entonces de la ecuación (19.1.3), se obtiene $\Pi_1 = 5$. Este resultado significa que el incremento de 1 dólar en la inversión (en el tiempo actual) producirá inmediatamente un incremento en el ingreso de 5 dólares, o sea, un aumento cinco veces mayor. En forma similar, bajo las condiciones supuestas, la ecuación (19.1.5) muestra que $\Pi_3 = 4$, lo cual significa que un incremento de 1 dólar en el gasto de inversión conllevará de inmediato a un incremento de 4 dólares en el gasto de consumo.

En el contexto de los modelos econométricos, ecuaciones tales como la (18.2.4) o $Q_t^d = Q_t^s$ (la cantidad demandada igual a la cantidad ofrecida) se conocen como *condiciones de equilibrio*. La identidad (18.2.4) establece que el ingreso agregado $Y$ debe ser igual al consumo agregado (es decir, gastos de consumo más gastos de inversión). Cuando se alcanza este equilibrio, las variables endógenas asumen sus valores de equilibrio.[3]

Observe una característica interesante de las ecuaciones en forma reducida. Puesto que solamente las variables predeterminadas y las perturbaciones estocásticas aparecen al lado derecho de estas ecuaciones, y puesto que se ha presumido que las variables predeterminadas no están correlacionadas con los términos de perturbación, el método de MCO puede aplicarse para estimar los coeficientes de las ecuaciones en forma reducida (las $\Pi$). A partir de éstos se pueden estimar los coeficientes estructurales (las $\beta$), como se muestra más adelante. Este procedimiento se conoce como **mínimos cuadrados indirectos** (MCI), y los coeficientes estructurales estimados se denominan estimaciones por MCI.

[2] En los modelos econométricos, las variables exógenas desempeñan un papel crucial. Con mucha frecuencia, esas variables están bajo el control directo del gobierno. Algunos ejemplos son la tasa de impuestos personales y empresariales, los subsidios y los seguros de desempleo, etcétera.

[3] Para mayores detalles, véase Jan Kmenta, Elements of Econometrics, 2a. ed., Macmillan, Nueva York, 1986, pp. 723-731.

En el capítulo 20 se estudiará el método de MCI con mayor detalle. Mientras tanto, observe que puesto que los coeficientes en la forma reducida pueden estimarse mediante el método de MCO, y puesto que dichos coeficientes son combinaciones de los coeficientes estructurales, existe la posibilidad de que estos últimos puedan ser "recuperados" a partir de los coeficientes en la forma reducida, y es en la estimación de los parámetros estructurales donde se puede estar finalmente interesado. ¿Cómo se recuperan los coeficientes estructurales a partir de los coeficientes en la forma reducida? La respuesta se da en la sección 19.2 y contiene la esencia del problema de identificación.

## 19.2 Problema de identificación

El **problema de identificación** pretende establecer si las estimaciones numéricas de los parámetros de una ecuación estructural pueden obtenerse de los coeficientes en forma reducida estimados. Si puede hacerse, se dice que la ecuación particular está *identificada*; si no, se dice entonces que la ecuación bajo consideración está *no identificada* o *subidentificada*.

Una ecuación identificada puede estar *exactamente* (o total o precisamente) *identificada* o *sobreidentificada*. Se dice que está exactamente identificada si pueden obtenerse valores numéricos únicos de los parámetros estructurales. Se dice que está sobreidentificada si puede obtenerse más de un valor numérico para algunos de los parámetros de las ecuaciones estructurales. Las circunstancias bajo las cuales puede ocurrir cada uno de los casos anteriores se indicarán en seguida.

El problema de identificación surge porque diferentes conjuntos de coeficientes estructurales pueden ser compatibles con el mismo conjunto de información. En otras palabras, una ecuación en una forma reducida dada puede ser compatible con diferentes ecuaciones estructurales o con diferentes hipótesis (modelos), y puede ser difícil decir cuál hipótesis (modelo) particular se está investigando. En lo que resta de la sección se consideran diversos ejemplos para mostrar la naturaleza del problema de identificación.

### Subidentificación

Considere de nuevo el modelo de demanda y oferta (18.2.1) y (18.2.2), conjuntamente con la condición de mercado nivelado, o de equilibrio, según la cual la demanda es igual a la oferta. Mediante la condición de equilibrio se obtiene:

$$\alpha_0 + \alpha_1 P_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.2.1}$$

Al resolver la ecuación (19.2.1) se obtiene el precio de equilibrio

$$P_t = \Pi_0 + v_t \tag{19.2.2}$$

en donde

$$\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \tag{19.2.3}$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \tag{19.2.4}$$

Al sustituir $P_t$ de la ecuación (19.2.2) en la ecuación (18.2.1) o (18.2.2), se obtiene la siguiente cantidad de equilibrio:

$$Q_t = \Pi_1 + w_t \tag{19.2.5}$$

en donde

$$\Pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \tag{19.2.6}$$

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \tag{19.2.7}$$

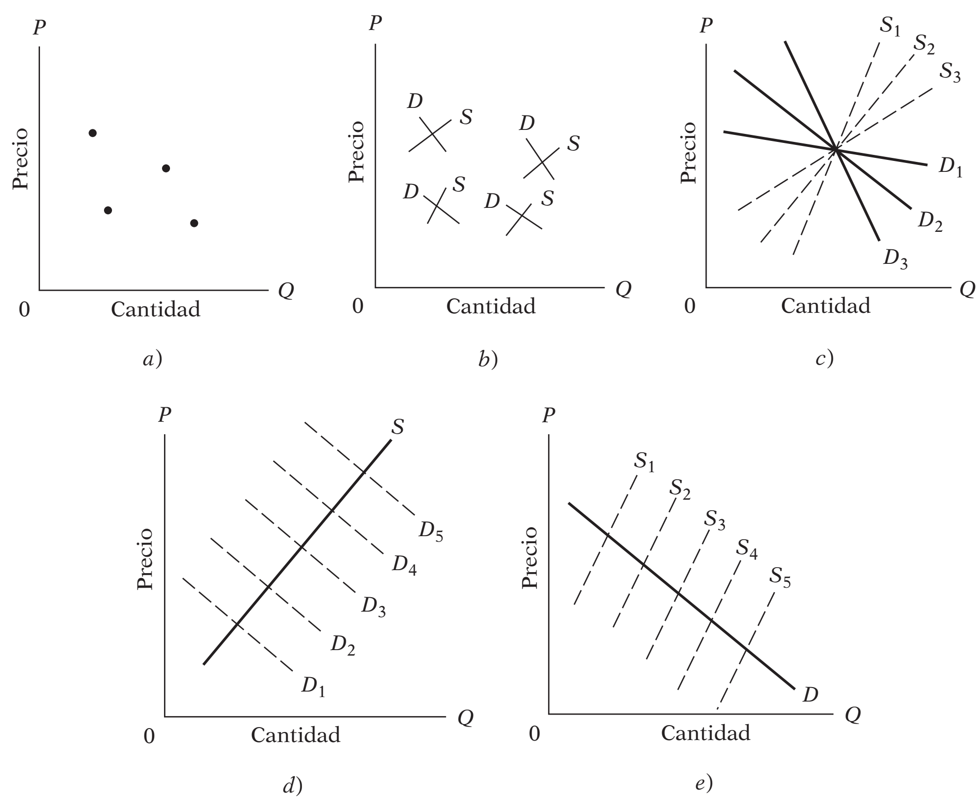
A propósito, es necesario observar que los términos de error $v_t$ y $w_t$ son combinaciones lineales de los términos de error originales $u_1$ y $u_2$.

Las ecuaciones (19.2.2) y (19.2.5) son ecuaciones en forma reducida. Ahora bien, el modelo de demanda y oferta contiene cuatro coeficientes estructurales $\alpha_0$, $\alpha_1$, $\beta_0$ y $\beta_1$, pero no hay una forma única de estimarlos. ¿Por qué? La respuesta se encuentra en los dos coeficientes en la forma reducida dados en las ecuaciones (19.2.3) y (19.2.6). Estos coeficientes contienen los cuatro parámetros estructurales, pero no hay forma de estimar las cuatro incógnitas estructurales a partir únicamente de dos coeficientes en forma reducida. En el álgebra de secundaria se decía que para estimar cuatro incógnitas se deben tener cuatro ecuaciones (independientes) y, en general, para estimar $k$ incógnitas se deben tener $k$ ecuaciones (independientes). A propósito, si se efectúa la regresión en forma reducida (19.2.2) y (19.2.5) se verá que no hay variables explicativas, sólo las constantes, y éstas simplemente darán los valores promedio de $P$ y $Q$ (¿por qué?).

Todo esto significa que, dada la información de series de tiempo sobre $P$ (precio) y $Q$ (cantidad) y ninguna información adicional, no hay forma de que el investigador pueda garantizar si está estimando la función de demanda o la función de oferta. Es decir, unos $P_t$ y $Q_t$ dados representan simplemente el punto de intersección de las curvas apropiadas de demanda y oferta en razón de la condición de equilibrio de que la demanda sea igual a la oferta. Para ver esto con claridad, considere el diagrama de dispersión que aparece en la figura 19.1.

La figura 19.1*a*) muestra algunos puntos dispersos que relacionan $Q$ con $P$. Cada punto disperso representa la intersección de una curva de demanda y de oferta, como se muestra en la figura 19.1*b*). Ahora considere un punto aislado como el que se muestra en la figura 19.1*c*). No hay forma de asegurar cuál de las curvas de demanda y oferta, entre toda la familia de curvas que aparece en ese panel, generó ese punto. Es claro que se requiere información adicional sobre la naturaleza de dichas curvas. Por ejemplo, si la curva de demanda se desplaza en el tiempo debido

**FIGURA 19.1**
Funciones hipotéticas de oferta y demanda y el problema de la identificación.



a cambios en el ingreso, en los gustos, etc., pero la de oferta permanece relativamente estable ante esos cambios, como se ilustra en la figura 19.1*d*), los puntos dispersos indican una curva de oferta. En esta situación, se dice que se ha identificado la curva de oferta. Mediante el mismo procedimiento, si la curva de oferta se desplaza en el tiempo debido a cambios en las condiciones climáticas (en el caso de bienes agrícolas) o debido a otros factores externos, pero la de demanda permanece relativamente estable, como se ilustra en la figura 19.1*e*), los puntos dispersos indican una curva de demanda. En este caso, se dice que la curva de demanda se ha identificado.

Hay una forma alterna y posiblemente más ilustrativa de considerar el problema de la identificación. Suponga que se multiplica la ecuación (18.2.1) por $\lambda(0 \leq \lambda \leq 1)$ y la ecuación (18.2.2) por $1 - \lambda$ para obtener las siguientes ecuaciones (*nota:* se eliminan aquí los superíndices de $Q$):

$$\lambda Q_t = \lambda\alpha_0 + \lambda\alpha_1 P_t + \lambda u_{1t} \tag{19.2.8}$$

$$(1-\lambda)Q_t = (1-\lambda)\beta_0 + (1-\lambda)\beta_1 P_t + (1-\lambda)u_{2t} \tag{19.2.9}$$

Al sumar estas dos ecuaciones, se obtiene la siguiente combinación lineal de las ecuaciones originales de demanda y oferta:

$$Q_t = \gamma_0 + \gamma_1 P_t + w_t \tag{19.2.10}$$

donde

$$\begin{aligned}\gamma_0 &= \lambda\alpha_0 + (1-\lambda)\beta_0 \\ \gamma_1 &= \lambda\alpha_1 + (1-\lambda)\beta_1 \\ w_t &= \lambda u_{1t} + (1-\lambda)u_{2t}\end{aligned} \tag{19.2.11}$$

La ecuación (19.2.10) "falsa" o "híbrida", a partir de la observación, no es distinguible de la ecuación (18.2.1) ni de la (18.2.2), porque éstas consideran las regresiones de $Q$ y de $P$. Por consiguiente, si se tiene información de series de tiempo sobre $P$ y $Q$ solamente, cualquiera de las ecuaciones (18.2.1), (18.2.2) o (19.2.10) puede ser compatible con la misma información. En otras palabras, la misma información puede ser compatible con la "hipótesis" de las ecuaciones (18.2.1), (18.2.2) o (19.2.10), y no hay forma de decir cuál de éstas se está verificando.

Para que una ecuación esté identificada, es decir, para que sus parámetros sean estimados, debe mostrarse que el conjunto dado de información no producirá una ecuación estructural que sea similar en apariencia a la ecuación en la cual se está interesado. Si se pretende estimar la función de demanda, se debe demostrar que la información dada no es consistente con la función de oferta ni con otro tipo de ecuación híbrida.

## Identificación precisa o exacta

La razón por la cual no fue posible identificar las anteriores funciones de demanda o de oferta fue porque las mismas variables $P$ y $Q$ están presentes en ambas funciones y no se dispone de información adicional, como la indicada en las figuras 19.1*d* o *e*. Pero suponga que se considera el siguiente modelo de demanda y oferta:

*Función de demanda*: $$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \alpha_1 < 0, \alpha_2 > 0 \tag{19.2.12}$$

*Función de oferta*: $$Q_t = \beta_0 + \beta_1 P_t + u_{2t} \qquad \beta_1 > 0 \tag{19.2.13}$$

donde $I$ = ingreso del consumidor, una variable exógena, y todas las demás variables como se definieron anteriormente.

Observe que la única diferencia entre el modelo anterior y el modelo original de demanda y oferta es que hay una variable adicional en la función de demanda, a saber, el ingreso. De la teoría económica de la demanda se sabe que el ingreso es, por lo general, un determinante importante de la demanda de la mayoría de bienes y servicios. Por consiguiente, su inclusión en la función de demanda proporcionará información adicional sobre el comportamiento del consumidor. Para la mayoría de los bienes se espera que el ingreso tenga un efecto positivo sobre el consumo ($\alpha_2 > 0$).

Al utilizar el mecanismo de nivelación del mercado, cantidad demandada = cantidad ofrecida, se tiene:

$$\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.2.14}$$

Al resolver la ecuación (19.2.14), se obtiene el siguiente valor de equilibrio de $P_t$:

$$P_t = \Pi_0 + \Pi_1 I_t + v_t \tag{19.2.15}$$

en donde los coeficientes en la forma reducida son

$$\begin{aligned} \Pi_0 &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \\ \Pi_1 &= -\frac{\alpha_2}{\alpha_1 - \beta_1} \end{aligned} \tag{19.2.16}$$

y

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

Al sustituir el valor de equilibrio de $P_t$ en la función de demanda u oferta anterior, se obtiene la siguiente cantidad de equilibrio:

$$Q_t = \Pi_2 + \Pi_3 I_t + w_t \tag{19.2.17}$$

en donde

$$\begin{aligned} \Pi_2 &= \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \\ \Pi_3 &= -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \end{aligned} \tag{19.2.18}$$

y

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

Puesto que las ecuaciones (19.2.15) y (19.2.17) son ecuaciones en forma reducida, puede aplicarse el método de MCO para estimar sus parámetros. Ahora bien, el modelo de demanda y oferta (19.2.12) y (19.2.13) contiene cinco coeficientes estructurales —$\alpha_0$, $\alpha_1$, $\alpha_2$, $\beta_0$, y $\beta_1$—, pero sólo se dispone de cuatro ecuaciones para estimarlos, a saber, los cuatro coeficientes en forma reducida $\Pi_0$, $\Pi_1$, $\Pi_2$ y $\Pi_3$, dados en las ecuaciones (19.2.16) y (19.2.18). Por tanto, no es posible encontrar una solución única para todos los coeficientes estructurales. Sin embargo, puede mostrarse con facilidad que los parámetros de la función de oferta pueden ser identificados (estimados) porque

$$\begin{aligned} \beta_0 &= \Pi_2 - \beta_1 \Pi_0 \\ \beta_1 &= \frac{\Pi_3}{\Pi_1} \end{aligned} \tag{19.2.19}$$

Pero no hay una forma única de estimar los parámetros de la función de demanda; por consiguiente, ésta permanece subidentificada. A propósito, observe que el coeficiente estructural $\beta_1$ es una función no lineal de los coeficientes en forma reducida, lo cual crea algunos problemas cuando se trata de estimar el error estándar del $\beta_1$ estimado, como se verá en el capítulo 20.

Para verificar que la función de demanda (19.2.12) no puede ser identificada (estimada), multiplique ésta por $\lambda$ $(0 \le \lambda \le 1)$ y la (19.2.13) por $1 - \lambda$, y luego sume para obtener la siguiente ecuación “híbrida”:

$$Q_t = \gamma_0 + \gamma_1 P_t + \gamma_2 I_t + w_t \tag{19.2.20}$$

en donde

$$\begin{aligned}\gamma_0 &= \lambda\alpha_0 + (1-\lambda)\beta_0\\ \gamma_1 &= \lambda\alpha_1 + (1-\lambda)\beta_1\\ \gamma_2 &= \lambda\alpha_2\end{aligned} \tag{19.2.21}$$

y

$$w_t = \lambda u_{1t} + (1-\lambda)u_{2t}$$

La ecuación (19.2.20) es, a partir de la observación, indistinguible de la función de demanda (19.2.12), aunque sí es distinguible de la función de oferta (19.2.13), que no contiene la variable *I* como una variable explicativa. Por tanto, la función de demanda permanece sin identificar.

**Observe un hecho interesante: ¡es la presencia de una variable adicional en la función de demanda la que permite identificar la función de oferta!** ¿Por qué? La inclusión de la variable ingreso en la ecuación de demanda proporciona alguna información adicional sobre la variabilidad de la función, como lo indica la figura 19.1*d*). La figura muestra cómo la intersección de la curva estable de oferta con la curva de demanda en movimiento (debido a cambios en el ingreso) permite trazar (identificar) la curva de oferta. Como se mostrará en breve, con mucha frecuencia la posibilidad de identificar una ecuación depende de si excluye una o más variables que están incluidas en otras ecuaciones del modelo.

Pero suponga que se considera el siguiente modelo de demanda y oferta:

*Función de demanda*: $$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \alpha_1 < 0, \alpha_2 > 0 \tag{19.2.12}$$

*Función de oferta*: $$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \qquad \beta_1 > 0, \beta_2 > 0 \tag{19.2.22}$$

donde la función de demanda permanece igual que antes pero la función de oferta incluye una variable explicativa adicional, el precio que está rezagado un periodo. La función de oferta postula que la cantidad de un bien ofrecido depende de su precio actual y del precio del periodo anterior, un modelo frecuentemente utilizado para explicar la oferta de muchos bienes agrícolas. Observe que $P_{t-1}$ es una variable predeterminada porque su valor se conoce en el tiempo *t*.

Por el mecanismo de nivelación del mercado se tiene que

$$\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \tag{19.2.23}$$

Al resolver esta ecuación se obtiene el siguiente precio de equilibrio:

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 P_{t-1} + v_t \tag{19.2.24}$$

en donde

$$\begin{aligned}\Pi_0 &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}\\ \Pi_1 &= -\frac{\alpha_2}{\alpha_1 - \beta_1}\\ \Pi_2 &= \frac{\beta_2}{\alpha_1 - \beta_1}\\ v_t &= \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}\end{aligned} \tag{19.2.25}$$

Al sustituir el precio de equilibrio en la ecuación de demanda o en la de oferta se obtiene la correspondiente cantidad de equilibrio:

$$Q_t = \Pi_3 + \Pi_4 I_t + \Pi_5 P_{t-1} + w_t \qquad \textbf{(19.2.26)}$$

en donde los coeficientes en la forma reducida son

$$\begin{aligned}\Pi_3 &= \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}\\ \Pi_4 &= -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}\\ \Pi_5 &= \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}\end{aligned} \qquad \textbf{(19.2.27)}$$

y

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

El modelo de demanda y oferta dado en las ecuaciones (19.2.12) y (19.2.22) contiene seis coeficientes estructurales —$\alpha_0$, $\alpha_1$, $\alpha_2$, $\beta_0$, $\beta_1$ y $\beta_2$—, y hay seis coeficientes en la forma reducida —$\Pi_0$, $\Pi_1$, $\Pi_2$, $\Pi_3$, $\Pi_4$ y $\Pi_5$— para estimarlos. Así, se tienen seis ecuaciones con seis incógnitas y normalmente es posible obtener estimaciones únicas. Por consiguiente, tanto los parámetros de ambas ecuaciones, de demanda y de oferta, como el sistema en su totalidad pueden ser identificados. (En el ejercicio 19.2 se pide al lector expresar los seis coeficientes estructurales en términos de los seis coeficientes en su forma reducida dados anteriormente, para mostrar que la estimación única del modelo es posible.)

Para verificar que las funciones de demanda y oferta anteriores son identificables, se puede recurrir también al mecanismo de multiplicar la ecuación de demanda (19.2.12) por $\lambda$ ($0 \leq \lambda \leq 1$) y la función de oferta (19.2.22) por $1 - \lambda$ y luego sumarlas para obtener una ecuación híbrida. Dicha ecuación tendrá las variables predeterminadas $I_t$ y $P_{t-1}$; por tanto, ésta será una ecuación por observación diferente tanto de la ecuación de demanda como también de la ecuación de oferta porque la primera no contiene a $P_{t-1}$ y la última no contiene a $I_t$.

## Sobreidentificación

Para ciertos bienes y servicios, el ingreso, al igual que la riqueza del consumidor, es un determinante importante de la demanda. Por consiguiente, al modificar la función de demanda (19.2.12) como se muestra a continuación, y manteniendo la función de oferta como antes, se obtiene:

*Función de demanda*: $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$ **(19.2.28)**

*Función de oferta*: $Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$ **(19.2.22)**

en donde, adicionalmente a las variables ya definidas, $R$ representa la riqueza; para la mayoría de los bienes y servicios se espera que la riqueza, al igual que el ingreso, tenga un efecto positivo sobre el consumo.

Al igualar la demanda a la oferta, se obtiene el siguiente precio y la siguiente cantidad de equilibrio:

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + \Pi_3 P_{t-1} + v_t \qquad \textbf{(19.2.29)}$$

$$Q_t = \Pi_4 + \Pi_5 I_t + \Pi_6 R_t + \Pi_7 P_{t-1} + w_t \qquad \textbf{(19.2.30)}$$

en donde

$$
\begin{aligned}
\Pi_0 &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} & \Pi_1 &= -\frac{\alpha_2}{\alpha_1 - \beta_1} \\
\Pi_2 &= -\frac{\alpha_3}{\alpha_1 - \beta_1} & \Pi_3 &= \frac{\beta_2}{\alpha_1 - \beta_1} \\
\Pi_4 &= \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} & \Pi_5 &= -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \\
\Pi_6 &= -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1} & \Pi_7 &= \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \\
w_t &= \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} & v_t &= \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}
\end{aligned}
\tag{19.2.31}
$$

El anterior modelo de demanda y oferta contiene siete coeficientes estructurales, pero hay ocho ecuaciones para estimarlos: los ocho coeficientes en forma reducida dados en (19.2.31); es decir, el número de ecuaciones es mayor que el número de incógnitas. Como resultado, no es posible obtener una estimación única de todos los parámetros del modelo, lo cual puede demostrarse fácilmente. De los anteriores coeficientes en la forma reducida se puede obtener

$$\beta_1 = \frac{\Pi_6}{\Pi_2} \tag{19.2.32}$$

o

$$\beta_1 = \frac{\Pi_5}{\Pi_1} \tag{19.2.33}$$

es decir, hay dos estimaciones del coeficiente de precios en la función de oferta y no hay garantía de que estos dos valores o soluciones sean idénticos.[4] Además, puesto que $\beta_1$ aparece en los denominadores de todos los coeficientes en su forma reducida, la ambigüedad en la estimación de $\beta_1$ será transmitida también a las demás estimaciones.

¿Por qué fue posible identificar la función de oferta en el sistema (19.2.12) y (19.2.22) pero no en el sistema (19.2.28) y (19.2.22), siendo que en ambos casos la función de oferta permanece igual? La respuesta es porque se tiene **demasiada información** para identificar la curva de oferta. Esta situación es la opuesta al caso de subidentificación, donde hay muy poca información. El exceso de información resulta del hecho de que en los modelos (19.2.12) y (19.2.22), la exclusión de la variable ingreso de la función de oferta fue suficiente para identificarla, pero en los modelos (19.2.28) y (19.2.22) la función de oferta excluye no solamente la variable ingreso, sino también la variable riqueza. En otras palabras, en el último modelo se impusieron "muchas" restricciones sobre la función de oferta al requerir excluir más variables de las necesarias para identificarla; sin embargo, esta situación no implica que la sobreidentificación necesariamente sea mala. En el capítulo 20 se verá la forma de manejar el problema de tener mucha información o muchas restricciones.

Ya se han examinado exhaustivamente todos los casos. Como muestra la exposición anterior, una ecuación en un modelo de ecuaciones simultáneas puede estar subidentificada o identificada (ya sea sobreidentificada o exactamente identificada). El modelo como un todo está identificado si cada una de sus ecuaciones también lo está. Para asegurar la identificación, se acude a las ecuaciones en forma reducida. En la sección 19.3 se considera un método alterno y posiblemente menos laborioso para determinar si una ecuación en un modelo de ecuaciones simultáneas está identificada o no.

[4] Observe la diferencia entre la subidentificación y la sobreidentificación. En el primer caso es imposible obtener estimaciones de los parámetros estructurales, en tanto que en el segundo puede haber varias estimaciones de uno o más coeficientes estructurales.

## 19.3 Reglas para la identificación

Como lo indican los ejemplos en la sección 19.2, en principio es posible recurrir a las ecuaciones en forma reducida para determinar la identificación de una ecuación en un sistema de ecuaciones simultáneas. Pero los ejemplos también muestran que este proceso puede llegar a ser muy dispendioso y laborioso. Por fortuna, no es indispensable utilizar este procedimiento. Las llamadas **condiciones de orden y de rango de identificación** aligeran la labor, proporcionando una rutina sistemática.

Para entender las condiciones de orden y de rango, se introduce la siguiente notación:

$M =$ número de variables endógenas en el modelo.
$m =$ número de variables endógenas en una ecuación dada.
$K =$ número de variables predeterminadas en el modelo, incluyendo el intercepto.
$k =$ número de variables predeterminadas en una ecuación dada.

### Condición de orden para la identificación[5]

Una condición necesaria (pero no suficiente) para la identificación, conocida como la **condición de orden**, puede expresarse en dos formas diferentes pero equivalentes, de la siguiente manera (las condiciones necesaria y suficiente para la identificación se presentan más adelante):

***Definición 19.1***

En un modelo de $M$ ecuaciones simultáneas, para que una ecuación esté identificada debe excluir *al menos* $M - 1$ variables (endógenas y predeterminadas) que aparecen en el modelo. Si excluye exactamente $M - 1$ variables, la ecuación está exactamente identificada. Si excluye más de $M - 1$ variables, estará sobreidentificada.

***Definición 19.2***

En un modelo de $M$ ecuaciones simultáneas, para que una ecuación esté identificada, el número de variables predeterminadas excluidas de esa ecuación no debe ser menor que el número de variables endógenas incluidas en la ecuación menos 1, es decir,

$$K - k \geq m - 1 \qquad \textbf{(19.3.1)}$$

Si $K - k = m - 1$, la ecuación está exactamente identificada, pero si $K - k > m - 1$, estará sobreidentificada.

En el ejercicio 19.1 se pide al lector demostrar que las dos definiciones anteriores son equivalentes.

Para ilustrar la condición de orden, considere de nuevo los ejemplos anteriores.

**EJEMPLO 19.1**

*Función de demanda:* $$Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t} \qquad \textbf{(18.2.1)}$$

*Función de oferta:* $$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t} \qquad \textbf{(18.2.2)}$$

Este modelo tiene dos variables endógenas $P$ y $Q$ y no tiene variables predeterminadas. Para ser identificadas cada una de estas ecuaciones debe excluir por lo menos $M - 1 = 1$ variable. Puesto que éste no es el caso, ninguna ecuación está identificada.

**EJEMPLO 19.2**

*Función de demanda:* $$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \textbf{(19.2.12)}$$

*Función de oferta:* $$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t} \qquad \textbf{(19.2.13)}$$

En este modelo, $Q$ y $P$ son endógenas e $I$ es exógena. Al aplicar la condición de orden dada en (19.3.1), se observa que la función de demanda no está identificada. Por otra parte, la función de oferta está exactamente identificada porque excluye de manera estricta $M - 1 = 1$ variable $I_t$.

[5] El término **orden** se refiere al orden de una matriz, es decir, el número de filas y de columnas que contiene. Véase el **apéndice B**.

**EJEMPLO 19.3**

*Función de demanda:* $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}$ **(19.2.12)**

*Función de oferta:* $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$ **(19.2.22)**

Dado que $P_t$ y $Q_t$ son endógenas e $I_t$ y $P_{t-1}$ son predeterminadas, la ecuación (19.2.12) excluye exactamente una variable $P_{t-1}$ y la ecuación (19.2.22) hace lo mismo con una variable $I_t$. Por tanto, cada ecuación está identificada según la condición de orden. Así, el modelo considerado como un todo también está identificado.

**EJEMPLO 19.4**

*Función de demanda:* $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$ **(19.2.28)**

*Función de oferta:* $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$ **(19.2.22)**

En este modelo $P_t$ y $Q_t$ son endógenas e $I_t$, $R_t$ y $P_{t-1}$ son predeterminadas. La función de demanda excluye exactamente una variable $P_{t-1}$ y, por tanto, según la condición de orden, está exactamente identificada. Pero la función de oferta excluye dos variables $I_t$ y $R_t$, y se dice entonces que está sobreidentificada. Como se mencionó, en este caso hay dos formas de estimar $\beta_1$, el coeficiente de la variable precio.

Aquí se observa una ligera complicación. Según la condición de orden, la función de demanda está identificada. Pero si se trata de estimar los parámetros de esta ecuación a partir de los coeficientes en forma reducida dados en (19.2.31), las estimaciones no serán únicas porque $\beta_1$, que forma parte de los cálculos, toma dos valores, y es preciso decidir cuál es el apropiado. Esta complicación puede obviarse porque, como se muestra en el capítulo 20, en casos de sobreidentificación el método de mínimos cuadrados indirectos no es apropiado y debe descartarse en favor de otros métodos. Uno de esos métodos es el de **mínimos cuadrados en dos etapas**, el cual se estudiará detalladamente en el capítulo 20.

Como lo indican los ejemplos anteriores, la identificación de una ecuación en un modelo de ecuaciones simultáneas es posible si dicha ecuación excluye una o más variables que están presentes en otras partes del modelo. Esta situación se conoce como *criterio de exclusión* (de variables), o *criterio de cero restricciones* (se supone que los coeficientes de las variables que no aparecen en una ecuación tienen valor de cero). Este criterio es el más utilizado para asegurar o determinar la identificación de una ecuación. Observe que el criterio de cero restricciones está basado en expectativas *a priori* o teóricas acerca de la ausencia de ciertas variables en una ecuación dada. Depende del investigador señalar claramente la razón por la cual espera que ciertas variables aparezcan en algunas ecuaciones y en otras no.

## Condición de rango para la identificación[6]

La condición de orden analizada anteriormente es una condición necesaria pero no suficiente para la identificación; es decir, aun si se cumple, puede suceder que una ecuación no esté identificada. Así, en el ejemplo 19.2, la ecuación de oferta fue identificada por la condición de orden porque excluyó la variable ingreso $I_t$, la cual aparecía en la función de demanda. Pero la identificación se logra solamente si $\alpha_2$, el coeficiente de $I_t$ en la función de demanda, no es cero, es decir, si la variable ingreso en verdad forma parte de la función de demanda, no sólo en forma probable, sino real.

En términos más generales, aun si una ecuación cumple la condición de orden $K - k \geq m - 1$, puede no estar identificada porque las variables predeterminadas excluidas de esa ecuación, pero presentes en el modelo, quizá no todas sean independientes de manera que tal vez no exista una correspondencia uno a uno entre los coeficientes estructurales (las $\beta$) y los coeficientes en forma

[6] El término **rango** se refiere al rango de una matriz y está dado por la matriz cuadrada de máximo rango (contenida en la matriz dada) cuyo determinante sea diferente de cero. De manera alterna, el rango de una matriz es el número máximo de filas o de columnas linealmente independientes de dicha matriz. Véase el **apéndice B**.

reducida (las Π). O sea, probablemente no sea posible estimar los parámetros estructurales a partir de los coeficientes en la forma reducida, como se mostrará en breve. Por consiguiente, se requiere una condición que sea tanto necesaria como suficiente para la identificación. Ésta es la condición de rango para la identificación, la cual puede expresarse en los siguientes términos:

***Condición de rango para la identificación***

En un modelo que contiene $M$ ecuaciones en $M$ variables endógenas, una ecuación está identificada si y sólo si puede construirse *por lo menos* un determinante diferente de cero, de orden $(M-1)(M-1)$, a partir de los coeficientes de las variables (endógenas y predeterminadas) excluidas de esa ecuación particular, pero incluidas en las otras ecuaciones del modelo.

Como ilustración de la condición de rango para la identificación, considere el siguiente sistema hipotético de ecuaciones simultáneas, en el cual las variables $Y$ son endógenas y las variables $X$ son predeterminadas.[7]

$$Y_{1t} - \beta_{10} \qquad - \beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} \qquad = u_{1t} \tag{19.3.2}$$

$$Y_{2t} - \beta_{20} \qquad - \beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} = u_{2t} \tag{19.3.3}$$

$$Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} \qquad - \gamma_{31}X_{1t} - \gamma_{32}X_{2t} = u_{3t} \tag{19.3.4}$$

$$Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} \qquad - \gamma_{43}X_{3t} = u_{4t} \tag{19.3.5}$$

Para facilitar la identificación, se escribe el sistema anterior en la tabla 19.1, que se explica por sí misma.

Primero se aplica la condición de orden para la identificación, como se muestra en la tabla 19.2. Cada ecuación está identificada por la condición de orden. Verifique esto con la condición de rango. Considere la primera ecuación, que excluye las variables $Y_4$, $X_2$ y $X_3$ (esta exclusión está representada por los ceros en el primer renglón de la tabla 19.1). Para que esta ecuación esté identificada, se debe obtener por lo menos un determinante diferente de cero de orden

**TABLA 19.1**

| | Coeficientes de las variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Núm. de ecuación** | **1** | **$Y_1$** | **$Y_2$** | **$Y_3$** | **$Y_4$** | **$X_1$** | **$X_2$** | **$X_3$** |
| (19.3.2) | $-\beta_{10}$ | 1 | $-\beta_{12}$ | $-\beta_{13}$ | 0 | $-\gamma_{11}$ | 0 | 0 |
| (19.3.3) | $-\beta_{20}$ | 0 | 1 | $-\beta_{23}$ | 0 | $-\gamma_{21}$ | $-\gamma_{22}$ | 0 |
| (19.3.4) | $-\beta_{30}$ | $-\beta_{31}$ | 0 | 1 | 0 | $-\gamma_{31}$ | $-\gamma_{32}$ | 0 |
| (19.3.5) | $-\beta_{40}$ | $-\beta_{41}$ | $-\beta_{42}$ | 0 | 1 | 0 | 0 | $-\gamma_{43}$ |

**TABLA 19.2**

| Núm. de ecuación | Núm. de variables predeterminadas excluidas $(K-k)$ | Núm. de variables endógenas incluidas menos uno, $(m-1)$ | ¿Identificadas? |
|---|---|---|---|
| (19.3.2) | 2 | 2 | Exactamente |
| (19.3.3) | 1 | 1 | Exactamente |
| (19.3.4) | 1 | 1 | Exactamente |
| (19.3.5) | 2 | 2 | Exactamente |

[7] El sistema de ecuaciones simultáneas expuesto en las ecuaciones (19.1.1) puede presentarse en la siguiente forma alterna, que puede ser conveniente para el manejo matricial.

$3 \times 3$, a partir de los coeficientes de las variables excluidas de esta ecuación, pero incluidas en otras. Para conseguir el determinante, se obtiene primero la matriz relevante de los coeficientes de las variables $Y_4$, $X_2$ y $X_3$ incluidas en las otras ecuaciones. En el presente caso, solamente hay una matriz como ésa, llamada **A**, definida de la siguiente manera:

$$\mathbf{A} = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix} \tag{19.3.6}$$

Se puede ver que el determinante de esta matriz es cero:

$$\det \mathbf{A} = \begin{vmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{vmatrix} \tag{19.3.7}$$

Puesto que el determinante es cero, el rango de la matriz (19.3.6), denotado por $\rho(\mathbf{A})$, es menor que 3. Por consiguiente, la ecuación (19.3.2) no satisface la condición de rango y, por tanto, no está identificada.

Como se anotó, la condición de rango es tanto necesaria como suficiente para la identificación. Por consiguiente, a pesar de que la condición de orden muestra que la ecuación (19.3.2) está identificada, la condición de rango muestra que no lo está. Al parecer, las columnas o los renglones de la matriz **A** dadas en (19.3.6) no son (linealmente) independientes, lo que significa que hay alguna relación entre las variables $Y_4, X_2$ y $X_3$. Como resultado, puede no haber suficiente información para estimar los parámetros de la ecuación (19.3.2); para el modelo anterior, las ecuaciones en forma reducida mostrarán que no es posible obtener los coeficientes estructurales de esa ecuación a partir de los coeficientes en la forma reducida. El lector debe verificar que, mediante la condición de rango, las ecuaciones (19.3.3) y (19.3.4) tampoco están identificadas, mientras que la ecuación (19.3.5) sí lo está.

Como lo muestra el análisis anterior, *la condición de rango dice si la ecuación bajo consideración está identificada o no, en tanto que la condición de orden expresa si dicha ecuación está exactamente identificada o sobreidentificada*.

Para aplicar la condición de rango, puede procederse de la siguiente manera:

1. Escriba el sistema en forma tabular, como aparece en la tabla 19.1.
2. Elimine los coeficientes del renglón en el cual aparece la ecuación bajo consideración.
3. Elimine también las columnas que corresponden a aquellos coeficientes del punto 2 que son diferentes de cero.
4. Los datos que quedan en la tabla corresponden únicamente a los coeficientes de las variables incluidas en el sistema pero no en la ecuación bajo consideración. Con estos datos, forme todas las matrices posibles, en este caso **A**, de orden $M - 1$ y obtenga los determinantes correspondientes. Si es posible encontrar al menos un determinante diferente de cero, la ecuación en cuestión estará identificada (en forma exacta o sobreidentificada). El rango de la matriz, por ejemplo **A**, en este caso, es exactamente igual a $M - 1$. Si todos los determinantes posibles $(M - 1)(M - 1)$ son cero, el rango de la matriz **A** es menor que $M - 1$ y la ecuación bajo investigación no está identificada.

El estudio de las condiciones de orden y de rango para la identificación conduce a los siguientes principios generales de identificabilidad de una ecuación estructural en un sistema de $M$ ecuaciones simultáneas:

1. Si $K - k > m - 1$ y el rango de matriz **A** es $M - 1$, la ecuación está sobreidentificada.
2. Si $K - k = m - 1$ y el rango de matriz **A** es $M - 1$, la ecuación está exactamente identificada.
3. Si $K - k \geq m - 1$ y el rango de la matriz **A** es menor que $M - 1$, la ecuación está subidentificada.
4. Si $K - k < m - 1$, la ecuación estructural no está identificada. El rango de la matriz **A** en este caso debe ser menor que M − 1. (¿Por qué?)

En adelante, cuando se hable de identificación, debe entenderse identificación exacta o sobreidentificación. No tiene sentido considerar ecuaciones no identificadas o subidentificadas puesto que, no importa qué tan completa sea la información, los parámetros estructurales no pueden ser estimados. Sin embargo, como se muestra en el capítulo 20, es posible identificar los parámetros de las ecuaciones sobreidentificadas al igual que aquellos de las ecuaciones exactamente identificadas.

¿Cuál condición se debe utilizar en la práctica: orden o rango? Para modelos grandes de ecuaciones simultáneas, la aplicación de la condición de rango es una labor muy dispendiosa. Por consiguiente, como afirma Harvey:

> Por fortuna, la condición de orden por lo general es suficiente para asegurar la identificación, y aunque es importante tener conciencia de la condición de rango, la no verificación de su cumplimiento raramente resultará en un desastre.[8]

## *19.4 Prueba de simultaneidad[9]

Si no hay ecuaciones simultáneas, o presencia del **problema de la simultaneidad**, MCO producen estimadores consistentes y eficientes. Por otra parte, si hay simultaneidad, los estimadores de MCO no son ni siquiera consistentes. Como se mostrará en el capítulo 20, en presencia de simultaneidad, los métodos de **mínimos cuadrados en dos etapas (MC2E)** y de **variables instrumentales (VI)** producirán estimadores consistentes y eficientes. Extrañamente, si se aplican métodos alternos cuando de hecho no hay simultaneidad, éstos producen estimadores que son consistentes pero no eficientes (es decir, con menor varianza). Todo este análisis sugiere que se debe verificar la presencia del problema de la simultaneidad antes de descartar los MCO en favor de las alternativas.

Como se mostró, el problema de la simultaneidad surge porque algunas de las regresoras son endógenas y, por consiguiente, es probable que estén correlacionadas con el término de perturbación o de error. Así, *en una prueba de simultaneidad, se intenta averiguar si una regresora (una endógena) está correlacionada con el término de error.* Si lo está, existe el problema de simultaneidad, en cuyo caso deben encontrarse alternativas a MCO; si no lo está, se pueden utilizar MCO. Para averiguar cuál es el caso en una situación concreta, se puede utilizar la prueba del error de especificación de Hausman.

### Prueba de especificación de Hausman

Una versión de esta prueba, que puede utilizarse para probar la presencia del problema de la simultaneidad, se explica de la siguiente manera.[10]

Para exponer la idea, considere el siguiente modelo de dos ecuaciones:

*Función de demanda:* $$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \tag{19.4.1}$$

*Función de oferta:* $$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.4.2}$$

en donde $P =$ precio
$Q =$ cantidad
$I =$ ingreso
$R =$ riqueza
$u =$ términos de error

Suponga que $I$ y $R$ son exógenas. Por supuesto, $P$ y $Q$ son endógenas.

---

* Opcional.

[8] Andrew Harvey, *The Econometric Analysis of Time Series*, 2a. ed., The MIT Press, Cambridge, Mass., 1990, p. 328.

[9] El siguiente análisis se tomó de Robert S. Pindyck y Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 3a. ed., McGraw-Hill, Nueva York, 1991, pp. 303-305.

[10] J.A. Hausman, "Specification Tests in Econometrics", *Econometrica*, vol. 46, noviembre de 1976, pp. 1251-1271. Véase también A. Nakamura y M. Nakamura, "On the Relationship among Several Specification Error Tests Presented by Durbin, Wu and Hausman", *Econometrica*, vol. 49, noviembre de 1981, pp. 1583-1588.

Ahora, considere la función de oferta (19.4.2). Si no hay problema de simultaneidad (es decir, $P$ y $Q$ son mutuamente independientes), entonces $P_t$ y $u_{2t}$ no deben estar correlacionadas (¿por qué?). Por otra parte, si hay simultaneidad, $P_t$ y $u_{2t}$ estarán correlacionadas. Para averiguar cuál es el caso, se realiza la siguiente prueba de Hausman:

Primero, de (19.4.1) y (19.4.2) se obtienen las siguientes ecuaciones en forma reducida:

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + v_t \tag{19.4.3}$$

$$Q_t = \Pi_3 + \Pi_4 I_t + \Pi_5 R_t + w_t \tag{19.4.4}$$

en donde $v$ y $w$ son los términos de error en la forma reducida. Si se estima (19.4.3) por MCO, se obtiene

$$\hat{P}_t = \hat{\Pi}_0 + \hat{\Pi}_1 I_t + \hat{\Pi}_2 R_t \tag{19.4.5}$$

Por consiguiente,

$$P_t = \hat{P}_t + \hat{v}_t \tag{19.4.6}$$

en donde $\hat{P}_t$ son estimaciones de $P_t$ y $\hat{v}_t$ son los residuos estimados. Al sustituir la ecuación (19.4.6) en la ecuación (19.4.2), se obtiene

$$Q_t = \beta_0 + \beta_1 \hat{P}_t + \beta_1 \hat{v}_t + u_{2t} \tag{19.4.7}$$

*Nota:* Los coeficientes de $P_t$ y de $v_t$ son los mismos. La diferencia entre esta ecuación y la ecuación original de oferta es que incluye la variable adicional $\hat{v}_t$, el residuo de la regresión (19.4.3).

Ahora, bajo la hipótesis nula de que no hay simultaneidad, la correlación entre $\hat{v}_t$ y $u_{2t}$ debe ser cero, asintóticamente. Así, si se efectúa la regresión (19.4.7) y se encuentra que el coeficiente de $v_t$ en la ecuación (19.4.7) es estadísticamente cero, puede concluirse que no hay problema de simultaneidad. Por supuesto, esta conclusión se invierte si se encuentra que este coeficiente sí es estadísticamente significativo. De paso, observe que la prueba de simultaneidad de Hausman se conoce también como la *prueba de endogeneidad de Hausman*. En el ejemplo anterior se quiere averiguar si $P_t$ es endógena. Si es así, se tiene el problema de la simultaneidad.

En esencia, la prueba de Hausman comprende los siguientes pasos:

**Paso 1.** Efectúe la regresión de $P_t$ sobre $I_t$ y $R_t$ para obtener $\hat{v}_t$.
**Paso 2.** Efectúe la regresión de $Q_t$ sobre $\hat{P}_t$ y $\hat{v}_t$ y realice una prueba $t$ sobre el coeficiente de $\hat{v}_t$. Si éste es significativo, no debe rechazar la hipótesis de simultaneidad; de otra forma, rechácela.[11] Sin embargo, para una estimación eficiente, Pindyck y Rubinfeld sugieren hacer la regresión de $Q_t$ sobre $P_t$ y $\hat{v}_t$.[12]

Existen otras formas de aplicar la prueba de Hausman, que se presentan por medio de un ejercicio.

**EJEMPLO 19.5**
*Modelo del gasto público de Pindyck-Rubinfeld*[13]

Para estudiar el comportamiento del gasto gubernamental de Estados Unidos a nivel estatal y local, los autores desarrollaron el siguiente modelo de ecuaciones simultáneas:

$$\text{EXP} = \beta_1 + \beta_2\text{AID} + \beta_3\text{INC} + \beta_4\text{POP} + u_i \tag{19.4.8}$$

$$\text{AID} = \delta_1 + \delta_2\text{EXP} + \delta_3\text{PS} + v_i \tag{19.4.9}$$

en donde EXP = gasto público de los gobiernos estatal y local
AID = nivel de ayuda mediante subsidio federal
INC = ingreso de los estados
POP = población estatal
PS = población estudiantil de primaria y secundaria
$u$ y $v$ = términos de error

En este modelo, INC, POP y PS se consideran exógenas.

[11] Si hay más de una regresora endógena, deberá utilizarse la prueba *F*.

[12] Pindyck y Rubinfeld, *op. cit.*, p. 304. *Nota:* La regresora es $P_t$ y *no* $\hat{P}_t$.

[13] Pindyck y Rubinfeld, *op. cit.*, pp. 176-177. La notación ha sido ligeramente alterada.

**EJEMPLO 19.5** (*continuación*)

Debido a la posibilidad de que existiera simultaneidad entre EXP y AID, los autores realizaron primero la regresión de AID sobre INC, POP y PS (es decir, la regresión en forma reducida). Sea $w_i$ el término de error en esta regresión y $\hat{w}_i$ el residuo calculado; después los autores efectuaron la regresión de EXP sobre AID, INC, POP y $\hat{w}_i$, para obtener los siguientes resultados:

$$\widehat{\text{EXP}} = -89.41 + 4.50\text{AID} + 0.00013\text{INC} - 0.518\text{POP} - 1.39\hat{w}_i$$

$$t = (-1.04) \quad (5.89) \quad (3.06) \quad (-4.63) \quad (-1.73) \qquad \textbf{(19.4.10)}^{14}$$

$$R^2 = 0.99$$

Al nivel de significancia de 5%, el coeficiente de $\hat{w}_i$ no es estadísticamente significativo y, por consiguiente, a este nivel no hay problema de simultaneidad. Sin embargo, al nivel de significancia de 10%, éste sí es estadísticamente significativo, y aumenta la posibilidad de que exista el problema de simultaneidad.

A propósito, la estimación MCO de la ecuación (19.4.8) es la siguiente:

$$\widehat{\text{EXP}} = -46.81 + 3.24\text{AID} + 0.00019\text{INC} - 0.597\text{POP}$$

$$t = (-0.56) \quad (13.64) \quad (8.12) \quad (-5.71) \qquad \textbf{(19.4.11)}$$

$$R^2 = 0.993$$

Se puede observar una característica interesante de los resultados dados en las ecuaciones (19.4.10) y (19.4.11): cuando la simultaneidad se considera de manera explícita, la variable *AID* es menos significativa, aunque numéricamente tiene una magnitud mayor.

## *19.5 Pruebas de exogeneidad

Como ya se dijo, es responsabilidad del investigador especificar cuáles variables son endógenas y cuáles exógenas. Esto dependerá del problema en cuestión y de la información *a priori* de la cual se disponga. Pero, ¿es posible desarrollar una prueba estadística de exogeneidad, al estilo de la prueba de causalidad de Granger?

La prueba de Hausman, analizada en la sección 19.4, puede utilizarse para responder a esta pregunta. Suponga que se tiene un modelo de tres ecuaciones con tres variables endógenas, $Y_1$, $Y_2$ y $Y_3$, y que hay tres variables exógenas, $X_1$, $X_2$ y $X_3$. Suponga además que la primera ecuación del modelo es

$$Y_{1i} = \beta_0 + \beta_2 Y_{2i} + \beta_3 Y_{3i} + \alpha_1 X_{1i} + u_{1i} \qquad \textbf{(19.5.1)}$$

Si $Y_2$ y $Y_3$ son verdaderamente endógenas, no se puede estimar la ecuación (19.5.1) por MCO (¿por qué?). Entonces, ¿cómo averiguarlo? Se puede proceder de la siguiente manera: se obtienen las ecuaciones en forma reducida para $Y_2$ y $Y_3$ (*Nota:* Las ecuaciones en forma reducida tendrán solamente variables predeterminadas al lado derecho). De estas ecuaciones se obtienen $\hat{Y}_{2i}$ y $\hat{Y}_{3i}$, los valores pronosticados de $Y_{2i}$ y $Y_{3i}$, respectivamente. Entonces, dentro del planteamiento de la prueba de Hausman, analizada anteriormente, se puede estimar la siguiente ecuación mediante MCO:

$$Y_{1i} = \beta_0 + \beta_2 Y_{2i} + \beta_3 Y_{3i} + \alpha_1 X_{1i} + \lambda_2 \hat{Y}_{2i} + \lambda_3 \hat{Y}_{3i} + u_{1i} \qquad \textbf{(19.5.2)}$$

Al emplear la prueba *F*, se demuestra la hipótesis de que $\lambda_2 = \lambda_3 = 0$. Si esta hipótesis es rechazada, $Y_2$ y $Y_3$ pueden considerarse endógenas, pero si no lo es, pueden ser tratadas como exógenas. Para un ejemplo concreto véase el ejercicio 19.16.

* Opcional.

[14] Al igual que en la nota 12 de pie de página, los autores utilizan a AID como regresora en lugar de AID.

## Resumen y conclusiones

1. El problema de la identificación es anterior al problema de la estimación.
2. Por problema de la identificación se entiende la posibilidad de obtener estimaciones numéricas únicas de los coeficientes estructurales a partir de los coeficientes en forma reducida.
3. Si esto puede hacerse, una ecuación que forma parte de un sistema de ecuaciones simultáneas está identificada. Si esto no puede hacerse, la ecuación estará no identificada o subidentificada.
4. Una ecuación identificada puede estarlo en forma exacta o estar sobreidentificada. En el primer caso pueden obtenerse valores únicos de los coeficientes estructurales; en el segundo, puede haber más de un valor para uno o más de los parámetros estructurales.
5. El problema de la identificación surge porque el mismo conjunto de información puede ser compatible con diferentes conjuntos de coeficientes estructurales, es decir, diferentes modelos. Así, en la regresión del precio sobre la cantidad solamente, es difícil decir si se está estimando la función de oferta o la de demanda, porque el precio y la cantidad forman parte de ambas ecuaciones.
6. Para establecer si una ecuación estructural está identificada, se puede aplicar la técnica de las **ecuaciones en forma reducida**, que expresan una variable endógena únicamente como función de variables predeterminadas.
7. Sin embargo, este laborioso procedimiento se puede evitar recurriendo a la **condición de orden** o a la **condición de rango** para la identificación. Aunque la condición de orden es fácil de aplicar, ésta proporciona solamente una condición necesaria para la identificación. Por otra parte, la condición de rango es una condición necesaria y suficiente para la identificación. Si la condición de rango se satisface, la de orden se satisface también, aunque lo contrario no es cierto. Pero, en la práctica, la condición de orden es generalmente adecuada para asegurar la identificabilidad.
8. En presencia de simultaneidad, por lo general, MCO no son aplicables, como se mostró en el capítulo 18. No obstante, si se desea utilizarlos es imperativo realizar explícitamente la prueba de simultaneidad. La **prueba de especificación de Hausman** puede emplearse para este propósito.
9. Aunque, en la práctica, la decisión de si una variable es endógena o exógena es un asunto de juicio, es posible utilizar la prueba de especificación de Hausman para determinar si una variable o un grupo de variables son exógenas o endógenas.
10. Aunque son de la misma familia, los conceptos de causalidad y de exogeneidad son diferentes y uno puede no necesariamente implicar el otro. En la práctica, es mejor mantener esos conceptos separados (véase la sección 17.14).

## EJERCICIOS

### Preguntas

19.1. Demuestre que las dos definiciones de la condición de orden para la identificación (véase la sección 19.3) son equivalentes.

19.2. Deduzca los coeficientes estructurales de los coeficientes en forma reducida dados en las ecuaciones (19.2.25) y (19.2.27).

19.3. Obtenga la forma reducida de los siguientes modelos y determine en cada caso si las ecuaciones estructurales no están identificadas, están identificadas exactamente o están sobreidentificadas:

*a*) Capítulo 18, ejemplo 18.2.
*b*) Capítulo 18, ejemplo 18.3.
*c*) Capítulo 18, ejemplo 18.6.

19.4. Verifique la identificabilidad de los modelos del ejercicio 19.3, aplicando las condiciones de orden y de rango para la identificación.

19.5. En el modelo (19.2.22) del texto se demostró que la ecuación de oferta estaba sobreidentificada. ¿Cuáles restricciones, de existir, sobre los parámetros estructurales harán que esta ecuación esté exactamente identificada? Justifique las restricciones que se impongan.

19.6. Del modelo

$$Y_{1t} = \beta_{10} + \beta_{12}Y_{2t} + \gamma_{11}X_{1t} + u_{1t}$$
$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + \gamma_{22}X_{2t} + u_{2t}$$

se obtienen las siguientes ecuaciones en forma reducida:

$$Y_{1t} = \Pi_{10} + \Pi_{11}X_{1t} + \Pi_{12}X_{2t} + w_t$$
$$Y_{2t} = \Pi_{20} + \Pi_{21}X_{1t} + \Pi_{22}X_{2t} + v_t$$

*a*) ¿Están identificadas las ecuaciones estructurales?

*b*) ¿Qué sucede con la identificación si se sabe *a priori* que $\gamma_{11} = 0$?

19.7. Remítase al ejercicio 19.6. Las ecuaciones estimadas en forma reducida son las siguientes:

$$Y_{1t} = 4 + 3X_{1t} + 8X_{2t}$$
$$Y_{2t} = 2 + 6X_{1t} + 10X_{2t}$$

*a*) Obtenga los valores de los parámetros estructurales.

*b*) ¿Cómo se probaría la hipótesis nula de que $\gamma_{11} = 0$?

19.8. El modelo

$$Y_{1t} = \beta_{10} + \beta_{12}Y_{2t} + \gamma_{11}X_{1t} + u_{1t}$$
$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + u_{2t}$$

genera las siguientes ecuaciones en forma reducida:

$$Y_{1t} = 4 + 8X_{1t}$$
$$Y_{2t} = 2 + 12X_{1t}$$

*a*) ¿Cuáles coeficientes estructurales, de existir, pueden ser estimados a partir de los coeficientes en forma reducida? Fundamente el argumento.

*b*) ¿Cómo cambia la respuesta de *a*) si se sabe *a priori* que 1) $\beta_{12} = 0$ y 2) $\beta_{10} = 0$?

19.9. Determine si las ecuaciones estructurales del modelo dado en el ejercicio 18.8 están identificadas.

19.10. Remítase al ejercicio 18.7 y determine cuáles ecuaciones estructurales pueden ser identificadas.

19.11. La tabla 19.3 es un modelo de cinco ecuaciones con cinco variables endógenas *Y* y cuatro variables exógenas *X*:

**TABLA 19.3**

| | Coeficientes de las variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Núm. de ecuación** | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ | $Y_5$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| 1 | 1 | $\beta_{12}$ | 0 | $\beta_{14}$ | 0 | $\gamma_{11}$ | 0 | 0 | $\gamma_{14}$ |
| 2 | 0 | 1 | $\beta_{23}$ | $\beta_{24}$ | 0 | 0 | $\gamma_{22}$ | $\gamma_{23}$ | 0 |
| 3 | $\beta_{31}$ | 0 | 1 | $\beta_{34}$ | $\beta_{35}$ | 0 | 0 | $\gamma_{33}$ | $\gamma_{34}$ |
| 4 | 0 | $\beta_{42}$ | 0 | 1 | 0 | $\gamma_{41}$ | 0 | $\gamma_{43}$ | 0 |
| 5 | $\beta_{51}$ | 0 | 0 | $\beta_{54}$ | 1 | 0 | $\gamma_{52}$ | $\gamma_{53}$ | 0 |

Determine la identificabilidad de cada ecuación con la ayuda de las condiciones de orden y de rango para la identificación.

19.12. Considere el siguiente modelo keynesiano ampliado de determinación del ingreso:

*Función de consumo:* $C_t = \beta_1 + \beta_2 Y_t - \beta_3 T_t + u_{1t}$

*Función de inversión:* $I_t = \alpha_0 + \alpha_1 Y_{t-1} + u_{2t}$

*Función de impuestos:* $T_t = \gamma_0 + \gamma_1 Y_t + u_{3t}$

*Identidad del ingreso:* $Y_t = C_t + I_t + G_t$

en donde $C =$ gasto de consumo
$Y =$ ingreso
$I =$ inversión
$T =$ impuestos
$G =$ gasto gubernamental
$u =$ términos de perturbación

En el modelo, las variables endógenas son $C, I, T$ y $Y$, y las variables predeterminadas son $G$ y $Y_{t-1}$.

Al aplicar la condición de orden, verifique la identificabilidad de cada una de las ecuaciones en el sistema y del sistema como un todo. ¿Qué sucedería si $r_t$, la tasa de interés, que se ha supuesto exógena, apareciera al lado derecho de la función de inversión?

19.13. Remítase a la información dada en la tabla 18.1 del capítulo 18. Utilizando esta información, estime las regresiones en forma reducida (19.1.2) y (19.1.4). ¿Se pueden estimar $\beta_0$ y $\beta_1$? Muestre los cálculos. ¿Está el modelo identificado? ¿Por qué sí o por qué no?

19.14. Suponga que se propone una definición adicional a la condición de orden para la identificación:

$$K \geq m + k - 1$$

que establece que el número de variables predeterminadas en el sistema no puede ser menor que el número de coeficientes desconocidos en la ecuación que va a ser identificada. Muestre que esta definición es equivalente a las otras dos definiciones de la condición de orden dadas en el texto.

19.15. A continuación se presenta una versión simplificada del modelo de Suits del mercado de sandías.*

*Ecuación de demanda:* $$P_t = \alpha_0 + \alpha_1(Q_t/N_t) + \alpha_2(Y_t/N_t) + \alpha_3 F_t + u_{1t}$$

*Función de oferta de la cosecha:* $$Q_t = \beta_0 + \beta_1(P_t/W_t) + \beta_2 P_{t-1} + \beta_3 C_{t-1} + \beta_4 T_{t-1} + u_{2t}$$

en donde $P =$ precio
$(Q/N) =$ cantidad demandada *per cápita*
$(Y/N) =$ ingreso *per cápita*
$F =$ costos de transporte
$(P/W) =$ precio relativo a la tasa salarial agrícola
$C =$ precio del algodón
$T =$ precio de otros vegetales
$N =$ población

$P$ y $Q$ son las variables endógenas.

*a*) Obtenga la forma reducida.

*b*) Determine si la función de demanda, de oferta, o las dos están identificadas.

## Ejercicios empíricos

19.16. Considere el siguiente modelo de demanda y oferta de dinero:

*Demanda de dinero:* $$M_t^d = \beta_0 + \beta_1 Y_t + \beta_2 R_t + \beta_3 P_t + u_{1t}$$

*Oferta de dinero:* $$M_t^s = \alpha_0 + \alpha_1 Y_t + u_{2t}$$

* D.B. Suits, "An Econometric Model of the Watermelon Market", *Journal of Farm Economics*, vol. 37, 1955, pp. 237-251.

**TABLA 19.4**
**Dinero, PIB, tasa de interés e índice de precios al consumidor en Estados Unidos, 1970-2006**

Fuente: *Economic Report of the President*, 2007, tablas B-2, B-60, B-69 y B-73.

| Observación | $M_2$ | PIB | TASABT | IPC |
|---|---|---|---|---|
| 1970 | 626.5 | 3 771.9 | 6.458 | 38.8 |
| 1971 | 710.3 | 3 898.6 | 4.348 | 40.5 |
| 1972 | 802.3 | 4 105.0 | 4.071 | 41.8 |
| 1973 | 855.5 | 4 341.5 | 7.041 | 44.4 |
| 1974 | 902.1 | 4 319.6 | 7.886 | 49.3 |
| 1975 | 1 016.2 | 4 311.2 | 5.838 | 53.8 |
| 1976 | 1 152.0 | 4 540.9 | 4.989 | 56.9 |
| 1977 | 1 270.3 | 4 750.5 | 5.265 | 60.6 |
| 1978 | 1 366.0 | 5 015.0 | 7.221 | 65.2 |
| 1979 | 1 473.7 | 5 173.4 | 10.041 | 72.6 |
| 1980 | 1 599.8 | 5 161.7 | 11.506 | 82.4 |
| 1981 | 1 755.5 | 5 291.7 | 14.029 | 90.9 |
| 1982 | 1 910.1 | 5 189.3 | 10.686 | 96.5 |
| 1983 | 2 126.4 | 5 423.8 | 8.63 | 99.6 |
| 1984 | 2 309.8 | 5 813.6 | 9.58 | 103.9 |
| 1985 | 2 495.5 | 6 053.7 | 7.48 | 107.6 |
| 1986 | 2 732.2 | 6 263.6 | 5.98 | 109.6 |
| 1987 | 2 831.3 | 6 475.1 | 5.82 | 113.6 |
| 1988 | 2 994.3 | 6 742.7 | 6.69 | 118.3 |
| 1989 | 3 158.3 | 6 981.4 | 8.12 | 124.0 |
| 1990 | 3 277.7 | 7 112.5 | 7.51 | 130.7 |
| 1991 | 3 378.3 | 7 100.5 | 5.42 | 136.2 |
| 1992 | 3 431.8 | 7 336.6 | 3.45 | 140.3 |
| 1993 | 3 482.5 | 7 532.7 | 3.02 | 144.5 |
| 1994 | 3 498.5 | 7 835.5 | 4.29 | 148.2 |
| 1995 | 3 641.7 | 8 031.7 | 5.51 | 152.4 |
| 1996 | 3 820.5 | 8 328.9 | 5.02 | 156.9 |
| 1997 | 4 035.0 | 8 703.5 | 5.07 | 160.5 |
| 1998 | 4 381.8 | 9 066.9 | 4.81 | 163.0 |
| 1999 | 4 639.2 | 9 470.3 | 4.66 | 166.6 |
| 2000 | 4 921.7 | 9 817.0 | 5.85 | 172.2 |
| 2001 | 5 433.5 | 9 890.7 | 3.45 | 177.1 |
| 2002 | 5 779.2 | 10 048.8 | 1.62 | 179.9 |
| 2003 | 6 071.2 | 10 301.0 | 1.02 | 184.0 |
| 2004 | 6 421.6 | 10 675.8 | 1.38 | 188.9 |
| 2005 | 6 691.7 | 11 003.4 | 3.16 | 195.3 |
| 2006 | 7 035.5 | 11 319.4 | 4.73 | 201.6 |

*Notas:* $M_2$ = oferta de dinero $M_2$ (miles de millones de dólares).
PIB = producto interno bruto (miles de millones de dólares).
TASABT = tasa de interés de los bonos del Tesoro a tres meses, %.
IPC = índice de precios al consumidor (1982-1984 = 100).

en donde $M$ = dinero
$Y$ = ingreso
$R$ = tasa de interés
$P$ = precio
$u$ = términos de error

Suponga que $R$ y $P$ son exógenas y que $M$ y $Y$ son endógenas. En la tabla 19.4 se presenta información sobre $M$ (definición $M_2$), $Y$ (*PIB*), $R$ (tasa de los bonos del Tesoro a tres meses) y $P$ (índice de precios al consumidor), para Estados Unidos durante 1970-2006.

*a*) ¿Está identificada la función de demanda?

*b*) ¿Está identificada la función de oferta?

*c*) Obtenga las expresiones para las ecuaciones en forma reducida para $M$ y para $Y$.

*d*) Aplique la prueba de la simultaneidad a la función de oferta.

*e*) ¿Cómo puede establecerse si $Y$ es realmente endógena en la función de oferta de dinero?

19.17. La prueba de Hausman analizada en el texto también se puede llevar a cabo de la siguiente manera. Considere la ecuación (19.4.7):

$$Q_t = \beta_0 + \beta_1 P_t + \beta_1 v_t + u_{2t}$$

*a*) Puesto que $P_t$ y $v_t$ tienen los mismos coeficientes, ¿cómo se probaría que en una aplicación determinada esto es así efectivamente? ¿Qué implicaciones tiene esto?

*b*) Puesto que $P_t$ no está correlacionada con $u_{2t}$ por definición (¿por qué?), una manera de averiguar si $P_t$ es exógena consiste en ver si $v_t$ está correlacionada con $u_{2t}$. ¿Cómo se puede probar esto? ¿Qué prueba se emplea? (*Pista:* Sustituya $P_t$ de la ecuación [19.4.6] en la ecuación [19.4.7].)

# Capítulo 20

# Métodos de ecuaciones simultáneas

Después de analizar la naturaleza de los modelos de ecuaciones simultáneas en los dos capítulos anteriores, aquí se trata el problema de la estimación de los parámetros de tales modelos. Para empezar, puede decirse que el problema de estimación es más bien complejo porque hay una diversidad de técnicas de estimación con propiedades estadísticas diversas. En vista de la naturaleza introductoria de este texto, se considerarán solamente algunas de estas técnicas. La exposición será simple y con frecuencia heurística, dejando los detalles más complejos para las referencias.

## 20.1 Enfoques para la estimación

Si se considera el modelo general de $M$ ecuaciones con $M$ variables endógenas dado en la ecuación (19.1.1), pueden adoptarse dos enfoques para estimar las ecuaciones estructurales, a saber: métodos uniecuacionales, también conocidos como **métodos de información limitada**, y métodos de sistemas, conocidos como **métodos de información completa**. En los métodos uniecuacionales que serán considerados en breve, cada ecuación en el sistema (de ecuaciones simultáneas) se estima individualmente, considerando las restricciones impuestas sobre ella (tales como la exclusión de algunas variables) sin preocuparse de las restricciones sobre las otras ecuaciones en el sistema;[1] de ahí el nombre de *métodos de información limitada*. Por otra parte, en los métodos de sistemas, se estiman todas las ecuaciones en el modelo de manera simultánea, teniendo en cuenta las restricciones ocasionadas por la omisión o ausencia de algunas variables sobre dichas ecuaciones (recuerde que tales restricciones son esenciales para la identificación), de aquí el nombre *métodos de información completa*.

Como ejemplo, considere el siguiente modelo de cuatro ecuaciones:

$$
\begin{aligned}
Y_{1t} &= \beta_{10} + \qquad\quad + \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \qquad\quad + \gamma_{11}X_{1t} + \qquad\qquad\qquad + u_{1t} \\
Y_{2t} &= \beta_{20} + \qquad\qquad\qquad\quad + \beta_{23}Y_{3t} \qquad\qquad + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} \qquad + u_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}Y_{1t} + \qquad\qquad\qquad + \beta_{34}Y_{4t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \quad + u_{3t} \\
Y_{4t} &= \beta_{40} + \qquad\quad + \beta_{42}Y_{2t} \qquad\qquad\qquad\qquad\qquad\qquad\quad + \gamma_{43}X_{3t} + u_{4t}
\end{aligned}
\tag{20.1.1}
$$

[1] Sin embargo, para fines de identificación, la información proporcionada por otras ecuaciones tendrá que ser considerada. Pero, como se mencionó en el capítulo 19, la estimación es posible sólo en el caso de ecuaciones identificadas (exactamente o sobreidentificadas). En este capítulo se supone que el problema de identificación se ha resuelto utilizando las técnicas de dicho capítulo.

en donde las $Y$ son las variables endógenas y las $X$ son las variables exógenas. Si se está interesado en estimar, por ejemplo, la tercera ecuación, los métodos uniecuacionales considerarán solamente esta ecuación, observando que las variables $Y_2$ y $X_3$ están excluidas de ella. En los métodos de sistemas, por el contrario, se trata de estimar las cuatro ecuaciones simultáneamente, teniendo en cuenta todas las restricciones impuestas sobre las diversas ecuaciones del sistema.

Para preservar el espíritu de los modelos de ecuaciones simultáneas, idealmente debería utilizarse el método de sistemas, tal como el **método de máxima verosimilitud con información completa (MVIC)**.[2] Sin embargo, en la práctica, tales métodos no son de uso frecuente por múltiples razones. Primero, la carga computacional es enorme. Por ejemplo, el modelo comparativamente pequeño (20 ecuaciones) de Klein-Goldberger de la economía de Estados Unidos para 1955 tenía 151 coeficientes diferentes de cero, de los cuales los autores estimaron sólo 51, utilizando información de series de tiempo. El modelo econométrico del Brookings Social Science Research Council (SSRC) para la economía estadounidense, publicado en 1965, tenía inicialmente 150 ecuaciones.[3] Aunque modelos tan elaborados pueden proporcionar detalles complejos de los diversos sectores de la economía, los cálculos representan un enorme esfuerzo aun en estos días de computadoras de alta velocidad, sin mencionar el costo involucrado. Segundo, los métodos de sistemas, como el MVIC, conducen a soluciones que son altamente no lineales en los parámetros y, por consiguiente, difíciles de determinar. Tercero, si hay un error de especificación (por ejemplo, una forma funcional equivocada o la exclusión de variables relevantes) en una o en más ecuaciones del sistema, dicho error es transmitido al resto del sistema. Como resultado, los métodos de sistemas se vuelven muy sensibles a los errores de especificación.

Por consiguiente, en la práctica, los métodos uniecuacionales son utilizados con más frecuencia. Como lo plantea Klein,

> Los métodos uniecuacionales, en el contexto de un sistema simultáneo, pueden ser menos sensibles a errores de especificación en el sentido de que aquellas partes del sistema que tienen una especificación correcta pueden no verse afectadas considerablemente por errores de especificación en otra parte.[4]

En el resto del capítulo se tratará sólo con los métodos uniecuacionales. En específico, se estudiarán los siguientes métodos uniecuacionales:

1. Mínimos cuadrados ordinarios (MCO)
2. Mínimos cuadrados indirectos (MCI)
3. Mínimos cuadrados en dos etapas (MC2E)

## 20.2 Modelos recursivos y mínimos cuadrados ordinarios

En el capítulo 18 se vio que, debido a la interdependencia entre el término de perturbación estocástico y la(s) variable(s) explicativa(s) endógena(s), el método de MCO es inapropiado para la estimación de una ecuación en un sistema de ecuaciones simultáneas. Así, como se explicó en la sección 18.3, si se aplica erróneamente, los estimadores no sólo resultan sesgados (en muestras pequeñas) sino también inconsistentes; es decir, sin importar qué tan grande sea el tamaño de la muestra, el sesgo no desaparece. Sin embargo, hay una situación en la cual el método de MCO puede ser aplicado apropiadamente, aun en el contexto de las ecuaciones simultáneas. Es el caso de los modelos **recursivos**, **triangulares** o **causales**.

[2] Para una exposición sencilla de este método, vea Carl F. Christ, *Econometric Models and Methods*, John Wiley & Sons, Nueva York, 1966, pp. 395-401.

[3] James S. Duesenberry, Gary Fromm, Lawrence R. Klein y Edwin Kuh, eds., *A Quarterly Model of the United States Economy*, Rand McNally, Chicago, 1965.

[4] Lawrence R. Klein, *A Textbook of Econometrics*, 2a. ed., Prentice-Hall, Englewood Cliffs, N.J., 1974, p. 150.

Para ver la naturaleza de estos modelos, considere el siguiente sistema de tres ecuaciones:

$$\begin{aligned} Y_{1t} &= \beta_{10} && + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \\ Y_{2t} &= \beta_{20} + \beta_{21}Y_{1t} && + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + u_{2t} \\ Y_{3t} &= \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} && + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + u_{3t} \end{aligned} \qquad \textbf{(20.2.1)}$$

en donde, como es usual, las $Y$ y las $X$ son las variables endógenas y exógenas respectivamente. Las perturbaciones son tales que

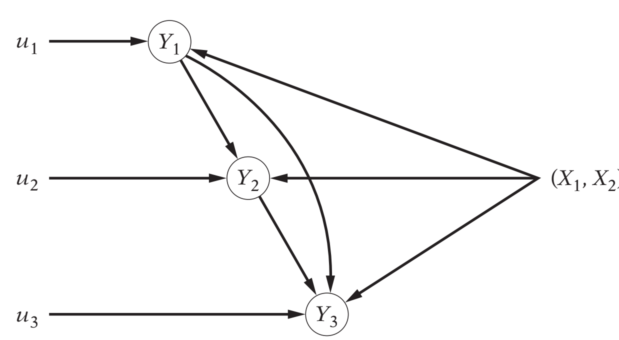
$$\text{cov}(u_{1t}, u_{2t}) = \text{cov}(u_{1t}, u_{3t}) = \text{cov}(u_{2t}, u_{3t}) = 0$$

es decir, las perturbaciones de diferentes ecuaciones en el mismo periodo no están correlacionadas (técnicamente, éste es el supuesto de **cero correlación contemporánea**).

Ahora, considere la primera ecuación de (20.2.1). Puesto que contiene variables exógenas al lado derecho y como, por los supuestos, no están correlacionadas con el término de perturbación $u_{1t}$, esta ecuación satisface el supuesto crítico del método de MCO clásico, a saber: la no correlación entre las variables explicativas y las perturbaciones estocásticas. Por tanto, MCO puede aplicarse directamente a esta ecuación. Luego, considere la segunda ecuación de (20.2.1), la cual contiene la variable endógena $Y_1$ como una variable explicativa junto con las $X$ no estocásticas. Ahora bien, MCO también puede ser aplicado a esta ecuación, siempre y cuando $Y_{1t}$ y $u_{2t}$ no estén correlacionadas. ¿Es esto así? La respuesta es sí porque $u_1$, el cual afecta a $Y_1$, por los supuestos y no está correlacionada con $u_2$. Por consiguiente, para todos los efectos prácticos, $Y_1$ es una variable predeterminada en lo que respecta a $Y_2$. Así, se puede proceder con la estimación de esta ecuación por MCO. Llevando este argumento un paso más adelante, también se puede aplicar MCO a la tercera ecuación en (20.2.1) porque $Y_1$ y $Y_2$ no están correlacionados con $u_3$.

Así, en el sistema recursivo, puede aplicarse MCO a cada ecuación en forma separada; de hecho, no se tiene el problema de las ecuaciones simultáneas en esta situación. Por la estructura de tales sistemas, es claro que no hay interdependencia entre las variables endógenas. Así, $Y_1$ afecta a $Y_2$ pero $Y_2$ no afecta a $Y_1$. En forma similar, $Y_1$ y $Y_2$ influyen en $Y_3$ sin que esta última las influya. En otras palabras, cada ecuación presenta una dependencia causal unilateral, de ahí el nombre de modelos causales.[5] La figura 20.1 ilustra esto de manera esquemática.

**FIGURA 20.1**
Modelo recursivo.




[5] El nombre alternativo *triangular* se desprende del hecho de que si se forma la matriz de los coeficientes de las variables endógenas dadas en (20.2.1), se obtiene la siguiente matriz triangular:

$$\begin{array}{l} \\ \text{Ecuación 1} \\ \text{Ecuación 2} \\ \text{Ecuación 3} \end{array} \begin{array}{c} \begin{matrix} Y_1 & Y_2 & Y_3 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ \beta_{31} & \beta_{32} & 1 \end{bmatrix} \end{array}$$

Observe que los datos por encima de la diagonal principal son cero (¿por qué?).

Como ejemplo de un sistema recursivo, puede postularse el siguiente modelo de determinación de salarios y precios:

$$\textit{Ecuación de precios:} \quad \dot{P}_t = \beta_{10} + \beta_{11}\dot{W}_{t-1} + \beta_{12}\dot{R}_t + \beta_{13}\dot{M}_t + \beta_{14}\dot{L}_t + u_{1t}$$
$$\textit{Ecuación de salarios:} \quad \dot{W}_t = \beta_{20} + \beta_{21}\text{UN}_t + \beta_{32}\dot{P}_t + u_{2t} \tag{20.2.2}$$

donde $\dot{P}$ = tasa de cambio del precio por unidad de producción
$\dot{W}$ = tasa de cambio de los salarios por empleado
$\dot{R}$ = tasa de cambio del precio del capital
$\dot{M}$ = tasa de cambio de los precios de importación
$\dot{L}$ = tasa de cambio de la productividad laboral
UN = tasa de desempleo, %[6]

La ecuación de precios postula que la tasa de cambio de los precios en el periodo actual es una función de las tasas de cambio en los precios del capital y de las materias primas, de la tasa de cambio en la productividad laboral y de la tasa de cambio en los salarios en el periodo anterior. La ecuación de salarios muestra que la tasa de cambio en los salarios en el periodo actual está determinada por la tasa de cambio de los precios en el periodo actual y por la tasa de desempleo. Es claro que la cadena causal va de $\dot{W}_{t-1} \to \dot{P}_t \to \dot{W}_t$ y, por consiguiente, puede aplicarse MCO para estimar los parámetros de las dos ecuaciones individualmente.

Aunque los modelos recursivos han demostrado ser útiles, la mayor parte de los modelos de ecuaciones simultáneas no presentan tal relación unilateral de causa y efecto. Por consiguiente, MCO, en general, resulta inapropiado para estimar una sola ecuación en el contexto de un modelo de ecuaciones simultáneas.[7]

Hay quienes sostienen que aunque MCO generalmente es inaplicable a modelos de ecuaciones simultáneas, se puede utilizar solamente como estándar o norma de comparación. Es decir, se puede estimar una ecuación estructural mediante MCO, con las propiedades resultantes de sesgo, inconsistencia, etc. Entonces, la misma ecuación puede ser estimada por otros métodos especialmente diseñados para manejar el problema de simultaneidad y los resultados de los dos métodos pueden compararse, por lo menos, de manera cualitativa. En muchas aplicaciones los resultados de MCO aplicado de forma inapropiada pueden no diferir mucho de aquellos obtenidos por métodos más complejos, como se verá más adelante. En principio, no debe haber mucha objeción en la presentación de resultados basados en MCO, siempre y cuando las estimaciones hechas con base en métodos alternos diseñados para modelos de ecuaciones simultáneas también sean proporcionadas. De hecho, este método podría dar alguna idea de qué tan malas son las estimaciones de MCO en situaciones en las cuales dicho método es aplicado inapropiadamente.[8]

---

[6] *Nota:* El símbolo con punto significa "derivada con respecto al tiempo". Por ejemplo, $\dot{P} + dP/dt$. Para series de tiempo discretas, $dP/dt$ en ocasiones es aproximada por $\Delta P/\Delta t$, donde el símbolo $\Delta$ es el operador de primeras diferencias, que fue introducido originalmente en el capítulo 12.

[7] Es importante tener presente que se está suponiendo que las perturbaciones entre ecuaciones no están correlacionadas contemporáneamente. Si éste no es el caso, puede ser que se deba recurrir a la técnica de estimación SURE (regresiones aparentemente no relacionadas) de Zellner para estimar los parámetros del sistema recursivo. Véase A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association*, vol. 57, 1962, pp. 348-368.

[8] Puede mencionarse que en muestras pequeñas, los estimadores alternativos, al igual que los estimadores por MCO, también son sesgados. Pero el estimador de MCO tiene la "virtud" de tener varianza mínima entre estos estimadores alternativos. Pero esto es cierto sólo para muestras pequeñas.

## 20.3 Estimación de una ecuación exactamente identificada: el método de mínimos cuadrados indirectos (MCI)

Para una ecuación estructural precisa o exactamente identificada, el método para obtener las estimaciones de los coeficientes estructurales a partir de las estimaciones por MCO de los coeficientes en forma reducida se conoce como **método de mínimos cuadrados indirectos (MCI),** y las estimaciones así obtenidas se conocen como **estimaciones de mínimos cuadrados indirectos**. MCI comprende los tres pasos siguientes:

**Paso 1.** Se obtienen primero las ecuaciones en forma reducida. Como se mencionó en el capítulo 19, éstas se obtienen de las ecuaciones estructurales en forma tal que la variable dependiente en cada ecuación es la única variable endógena y está en función únicamente de las variables predeterminadas (exógenas o endógenas rezagadas) y del (los) término(s) de error(es) estocástico(s).

**Paso 2.** Se aplica MCO individualmente a las ecuaciones en la forma reducida. Esta operación es permisible puesto que las variables explicativas en estas ecuaciones están predeterminadas y, por tanto, no están correlacionadas con las perturbaciones estocásticas. Las estimaciones así obtenidas son consistentes.[9]

**Paso 3.** Se obtienen estimaciones de los coeficientes estructurales originales a partir de los coeficientes en forma reducida estimados, obtenidos en el paso 2. Como se mencionó en el capítulo 19, si una ecuación está exactamente identificada, hay una correspondencia uno a uno entre los coeficientes estructurales y los coeficientes en la forma reducida; es decir, pueden derivarse estimaciones únicas de los primeros a partir de los últimos.

Como lo indica este procedimiento de tres etapas, el nombre de MCI se deriva del hecho de que los coeficientes estructurales (objeto principal de investigación en la mayoría de los casos) se obtienen indirectamente a partir de las estimaciones por MCO de los coeficientes en forma reducida.

### Ejemplo ilustrativo

Considere el modelo de demanda y oferta presentado en la sección 19.2, el cual, por conveniencia, está dado en seguida con un ligero cambio en la notación:

*Función de demanda:* $$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_{1t} \quad \textbf{(20.3.1)}$$

*Función de oferta:* $$Q_t = \beta_0 + \beta_1 P_t + u_{2t} \quad \textbf{(20.3.2)}$$

en donde $Q =$ cantidad
$P =$ precio
$X =$ ingreso o gasto

Suponga que $X$ es exógena. Como se mencionó antes, la función de oferta está exactamente identificada, en tanto que la función de demanda no lo está.

Las ecuaciones en forma reducida correspondientes a las ecuaciones estructurales anteriores son:

$$P_t = \Pi_0 + \Pi_1 X_t + w_t \quad \textbf{(20.3.3)}$$

$$Q_t = \Pi_2 + \Pi_3 X_t + v_t \quad \textbf{(20.3.4)}$$

[9] Además de ser consistentes, las estimaciones "pueden ser las mejores insesgadas y/o asintóticamente eficientes respectivamente, dependiendo de que: (*i*) las *z* [= *X*] sean exógenas y no simplemente predeterminadas [es decir, que no contengan valores rezagados de variables endógenas] y/o (*ii*) la distribución de las perturbaciones sea normal". (W.C. Hood y Tjalling C. Koopmans, *Studies in Econometric Method*, John Wiley & Sons, Nueva York, 1953, p. 133.)

en donde las $\Pi$ son los coeficientes en la forma reducida y son combinaciones (no lineales) de los coeficientes estructurales, como se muestra en las ecuaciones (19.2.16) y (19.2.18), y donde $w$ y $v$ son combinaciones lineales de las perturbaciones estructurales $u_1$ y $u_2$.

Observe que cada ecuación en la forma reducida contiene una sola variable endógena, que es la dependiente y está en función únicamente de la variable exógena $X$ (el ingreso) y de las perturbaciones estocásticas. Por tanto, los parámetros de las ecuaciones en la forma reducida anteriores pueden ser estimados por MCO. Las estimaciones son:

$$\hat{\Pi}_1 = \frac{\sum p_t x_t}{\sum x_t^2} \qquad \textbf{(20.3.5)}$$

$$\hat{\Pi}_0 = \bar{P} - \hat{\Pi}_1 \bar{X} \qquad \textbf{(20.3.6)}$$

$$\hat{\Pi}_3 = \frac{\sum q_t x_t}{\sum x_t^2} \qquad \textbf{(20.3.7)}$$

$$\hat{\Pi}_2 = \bar{Q} - \hat{\Pi}_3 \bar{X} \qquad \textbf{(20.3.8)}$$

en donde las letras en minúsculas, como es usual, denotan desviaciones de las medias muestrales, y $\bar{Q}$ y $\bar{P}$ son los valores de la media muestral de $Q$ y de $P$. Como se indicó antes, las $\Pi_i$ son estimadores consistentes y, bajo los supuestos apropiados, también son insesgados, con varianza mínima o asintóticamente eficientes (véase nota 9 de pie de página).

Puesto que el objetivo principal es determinar los coeficientes estructurales, se verá si es posible estimarlos a partir de los coeficientes en forma reducida. Ahora, como se indicó en la sección 19.2, la función de oferta está exactamente identificada. Por consiguiente, sus parámetros pueden estimarse de manera única a partir de los coeficientes en forma reducida del siguiente modo:

$$\beta_0 = \Pi_2 - \beta_1 \Pi_0 \qquad \text{y} \qquad \beta_1 = \frac{\Pi_3}{\Pi_1}$$

Por tanto, las estimaciones de estos parámetros pueden obtenerse a partir de las estimaciones de los coeficientes en forma reducida como

$$\hat{\beta}_0 = \hat{\Pi}_2 - \hat{\beta}_1 \hat{\Pi}_0 \qquad \textbf{(20.3.9)}$$

$$\hat{\beta}_1 = \frac{\hat{\Pi}_3}{\hat{\Pi}_1} \qquad \textbf{(20.3.10)}$$

que son los estimadores por MCI. Observe que los parámetros de la función de demanda no pueden ser estimados de esta forma (sin embargo, véase el ejercicio 20.13).

Para dar algunos resultados numéricos, se obtuvo la información que aparece en la tabla 20.1. Primero se estiman las ecuaciones en la forma reducida, efectuando separadamente las regresiones del precio y de la cantidad sobre el gasto de consumo real per cápita. Los resultados son los siguientes:

$$\begin{aligned} \hat{P}_t &= \ 90.9601 \ + \ 0.0007X_t \\ \text{ee} &= \ (4.0517) \quad (0.0002) \\ t &= (22.4499) \quad (3.0060) \qquad R^2 = (0.2440) \end{aligned} \qquad \textbf{(20.3.11)}$$

$$\begin{aligned} \hat{Q}_t &= \ 59.7618 \ + \ 0.0020X_t \\ \text{ee} &= \ (1.5600) \quad (0.00009) \\ t &= (38.3080) \quad (20.9273) \qquad R^2 = 0.9399 \end{aligned} \qquad \textbf{(20.3.12)}$$

Al utilizar las ecuaciones (20.3.9) y (20.3.10), se obtienen estas estimaciones por MCI:

$$\hat{\beta}_0 = -183.7043 \qquad \textbf{(20.3.13)}$$

$$\hat{\beta}_1 = \ 2.6766 \qquad \textbf{(20.3.14)}$$

**TABLA 20.1**
**Producción de cultivos, precios de cultivos y gasto de consumo personal *per cápita*, en dólares de 2007; Estados Unidos, 1975-2004**

Fuente: *Economic Report of the President*, 2007. Datos sobre $Q$ (tabla B-99), sobre $P$ (tabla B-101) y sobre $X$ (tabla B-31).

| Observación | Índice de producción de cultivos (1996 = 100), $Q$ | Índice de precios de cultivos que se pagan a los agricultores (1990-1992 = 100), $P$ | Gasto de consumo personal real *per cápita*, $X$ |
|---|---|---|---|
| 1975 | 66 | 88 | 4 789 |
| 1976 | 67 | 87 | 5 282 |
| 1977 | 71 | 83 | 5 804 |
| 1978 | 73 | 89 | 6 417 |
| 1979 | 78 | 98 | 7 073 |
| 1980 | 75 | 107 | 7 716 |
| 1981 | 81 | 111 | 8 439 |
| 1982 | 82 | 98 | 8 945 |
| 1983 | 71 | 108 | 9 775 |
| 1984 | 81 | 111 | 10 589 |
| 1985 | 85 | 98 | 11 406 |
| 1986 | 82 | 87 | 12 048 |
| 1987 | 84 | 86 | 12 766 |
| 1988 | 80 | 104 | 13 685 |
| 1989 | 86 | 109 | 14 546 |
| 1990 | 90 | 103 | 15 349 |
| 1991 | 90 | 101 | 15 722 |
| 1992 | 96 | 101 | 16 485 |
| 1993 | 91 | 102 | 17 204 |
| 1994 | 101 | 105 | 18 004 |
| 1995 | 96 | 112 | 18 665 |
| 1996 | 100 | 127 | 19 490 |
| 1997 | 104 | 115 | 20 323 |
| 1998 | 105 | 107 | 21 291 |
| 1999 | 108 | 97 | 22 491 |
| 2000 | 108 | 96 | 23 862 |
| 2001 | 108 | 99 | 24 722 |
| 2002 | 107 | 105 | 25 501 |
| 2003 | 108 | 111 | 26 463 |
| 2004 | 112 | 117 | 27 937 |

Por consiguiente, la regresión estimada por MCI es[10]

$$\hat{Q}_t = -183.7043 + 2.6766P_t \tag{20.3.15}$$

Para comparar, se dan los resultados de la regresión por MCO (aplicado de manera inapropiada) de $Q$ sobre $P$:

$$\begin{aligned} \hat{Q}_t &= \;\; 20.89 + \;\; 0.673P_t \\ \text{ee} &= (23.04) \quad (0.2246) \\ t &= \;\; (0.91) \quad (2.99) \qquad R^2 = 0.2430 \end{aligned} \tag{20.3.16}$$

Estos resultados muestran cómo MCO puede distorsionar el "verdadero" panorama cuando se aplica en situaciones inapropiadas.

[10] No se han presentado los errores estándar de los coeficientes estructurales estimados porque, como ya se mencionó, estos coeficientes generalmente son funciones no lineales de los coeficientes en la forma reducida y no hay un método simple de estimar sus errores estándar a partir de los errores estándar de los coeficientes en la forma reducida. Para tamaños de muestra grandes, sin embargo, los errores estándar de los coeficientes estructurales pueden ser obtenidos aproximadamente. Para mayores detalles, véase Jan Kmenta, *Elements of Econometrics*, Macmillan, Nueva York, 1971, p. 444.

### Propiedades de los estimadores por MCI

Se ha visto que los estimadores de los coeficientes en forma reducida son consistentes y, bajo los supuestos apropiados, también los mejores insesgados o asintóticamente eficientes (véase nota 9 de pie de página). ¿Se trasladan estas propiedades a los estimadores por MCI? Puede demostrarse que los estimadores por MCI heredan todas las propiedades asintóticas de los estimadores en forma reducida, tales como consistencia y eficiencia asintótica. Pero propiedades (de muestras pequeñas) tales como insesgamiento generalmente no continúan siendo válidas. En el apéndice 20A, sección 20A.1, se demuestra que los estimadores por MCI $\hat{\beta}_0$ y $\hat{\beta}_1$ de la función de oferta dados anteriormente son sesgados, pero el sesgo desaparece a medida que el tamaño de la muestra aumenta indefinidamente (es decir, los estimadores son consistentes).[11]

## 20.4 Estimación de una ecuación sobreidentificada: método de mínimos cuadrados en dos etapas (MC2E)

Considere el siguiente modelo:

*Función de ingreso:*
$$Y_{1t} = \beta_{10} + \qquad + \beta_{11}Y_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \tag{20.4.1}$$

*Función de oferta monetaria:*
$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} \qquad + u_{2t} \tag{20.4.2}$$

en donde $Y_1$ = ingreso
$Y_2$ = existencias de dinero
$X_1$ = gasto de inversión
$X_2$ = gasto del gobierno en bienes y servicios

Las variables $X_1$ y $X_2$ son exógenas.

La ecuación de ingreso, un híbrido de los enfoques de las teorías cuantitativa y keynesiana de la determinación del ingreso, establece que el ingreso está determinado por la oferta monetaria, el gasto de inversión y el gasto del gobierno. La *función de la oferta monetaria* postula que las existencias de dinero están determinadas (por el Sistema de la Reserva Federal) con base en el nivel del ingreso. Como es obvio, se tiene un problema de ecuaciones simultáneas, el cual puede verificarse mediante la prueba de simultaneidad estudiada en el capítulo 19.

Al aplicar la condición de orden para la identificación, puede verse que la ecuación del ingreso está subidentificada, en tanto que la ecuación de la oferta monetaria está sobreidentificada. Es poco lo que puede hacerse sobre la ecuación del ingreso, a no ser que se altere la especificación del modelo. La función de la oferta monetaria sobreidentificada no puede estimarse mediante MCI porque hay dos estimaciones de $\beta_{21}$ (el lector debe verificar esto por medio de los coeficientes en la forma reducida).

A manera de práctica, puede aplicarse MCO a la ecuación de la oferta monetaria, pero las estimaciones obtenidas por este mecanismo serán inconsistentes en vista de la probable correlación entre la variable explicativa estocástica $Y_1$ y el término de perturbación estocástico $u_2$. Suponga, sin embargo, que se encuentra una "variable representante" para la variable explicativa estocástica $Y_1$, tal que, aunque "se parece" a $Y_1$ (en el sentido de que está altamente correlacionada con $Y_1$), no está correlacionada con $u_2$. Tal variable se conoce también como **variable instrumental** (véase el capítulo 17). Si se puede encontrar tal variable representante, puede utilizarse MCO directamente para estimar la función de oferta monetaria. Pero ¿cómo se obtiene

[11] Intuitivamente esto puede verse de la siguiente manera: $E(\hat{\beta}_1) = \beta_1$ si $E(\hat{\Pi}_3/\hat{\Pi}_1) = (\Pi_3/\Pi_1)$. Ahora bien, aun si $E(\hat{\Pi}_3) = \Pi_3$ y $E(\hat{\Pi}_1) = \Pi_1$, puede demostrarse que $E(\hat{\Pi}_3/\hat{\Pi}_1) \neq E(\Pi_3)/E(\Pi_1)$; es decir, la esperanza de la razón de dos variables no es igual a la razón de las esperanzas de las dos variables. Sin embargo, como se muestra en el apéndice 20A.1, $\text{plím}(\hat{\Pi}_3/\hat{\Pi}_1) = \text{plím}(\hat{\Pi}_3)/\text{plím}(\hat{\Pi}_1) = \Pi_3/\Pi_1)$, puesto que $\hat{\Pi}_3$ y $\hat{\Pi}_1$ son estimadores consistentes.

semejante variable instrumental? Una respuesta está dada por el método de **mínimos cuadrados en dos etapas** (MC2E), desarrollado independientemente por Henri Theil[12] y Robert Basmann.[13] Como lo indica el nombre, el método comprende dos aplicaciones sucesivas de MCO. El proceso es el siguiente:

**Etapa 1.** Para eliminar la correlación probable entre $Y_1$ y $u_2$, efectúe primero la regresión de $Y_1$ sobre *todas* las variables predeterminadas en el *sistema completo*, no solamente en esa ecuación. En el presente caso, esto significa efectuar la regresión de $Y_1$, sobre $X_1$ y $X_2$ de la siguiente manera:

$$Y_{1t} = \hat{\Pi}_0 + \hat{\Pi}_1 X_{1t} + \hat{\Pi}_2 X_{2t} + \hat{u}_t \tag{20.4.3}$$

donde $\hat{u}_t$ son los residuos de MCO tradicionales. De la ecuación (20.4.3) se obtiene

$$\hat{Y}_{1t} = \hat{\Pi}_0 + \hat{\Pi}_1 X_{1t} + \hat{\Pi}_2 X_{2t} \tag{20.4.4}$$

donde $\hat{Y}_{1t}$ es una estimación del valor medio de $Y$ condicional de las $X$ fijas. Observe que (20.4.3) no es otra cosa que una regresión en forma reducida porque solamente las variables exógenas o predeterminadas aparecen en el lado derecho.

La ecuación (20.4.3) puede expresarse ahora como

$$Y_{1t} = \hat{Y}_{1t} + \hat{u}_t \tag{20.4.5}$$

lo cual muestra que la $Y_1$ estocástica consta de dos partes: $\hat{Y}_{1t}$, que es una combinación lineal de las $X$ no estocásticas, y un componente aleatorio $\hat{u}_t$. Siguiendo la teoría de MCO, $\hat{Y}_{1t}$ y $\hat{u}_t$ no están correlacionadas. (¿Por qué?)

**Etapa 2.** La ecuación sobreidentificada de la oferta monetaria puede escribirse ahora como

$$\begin{aligned} Y_{2t} &= \beta_{20} + \beta_{21}(\hat{Y}_{1t} + \hat{u}_t) + u_{2t} \\ &= \beta_{20} + \beta_{21}\hat{Y}_{1t} + (u_{2t} + \beta_{21}\hat{u}_t) \\ &= \beta_{20} + \beta_{21}\hat{Y}_{1t} + u_t^* \end{aligned} \tag{20.4.6}$$

en donde $u_t^* = u_{2t} + \beta_{21}\hat{u}_t$.

Al comparar la ecuación (20.4.6) con la ecuación (20.4.2), se observa que son muy similares en apariencia, siendo la única diferencia que $Y_1$ ha sido reemplazada por $\hat{Y}_1$. ¿Cuál es la ventaja de la ecuación (20.4.6)? Puede mostrarse que aunque en la ecuación original de oferta monetaria $Y_1$ está correlacionada o es probable que esté correlacionada con el término de perturbación $u_2$ (de aquí resulta que MCO sea inapropiado), $\hat{Y}_{1t}$ en la ecuación (20.4.6) no está correlacionada con $u_t^*$ *asintóticamente*, es decir, en muestras grandes (o en forma más precisa, a medida que el tamaño de la muestra aumenta indefinidamente). Como resultado, puede aplicarse MCO a la ecuación (20.4.6), lo cual dará estimaciones consistentes de los parámetros de la función de oferta de dinero.[14]

[12] Henri Theil, "Repeated Least-Squares Applied to Complete Equation Systems", The Hague Oficina Central de Planeación, Países Bajos, 1953 (mimeografiada).

[13] Robert L. Basmann, "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation", *Econometrica*, vol. 25, 1957, pp. 77-83.

[14] Observe que en muestras pequeñas es probable que $\hat{Y}_{1t}$ esté correlacionada con $u_i^*$. La razón es la siguiente: de la ecuación (20.4.4) se ve que $\hat{Y}_{1t}$ es una combinación lineal ponderada de las $X$ predeterminadas con las $\hat{\Pi}$ como ponderaciones. Ahora, aun si las variables predeterminadas son verdaderamente no estocásticas, las $\hat{\Pi}$, siendo estimadores, son estocásticas. Por consiguiente, $\hat{Y}_{1t}$ también es estocástica. Ahora bien, del análisis de las ecuaciones en la forma reducida y de la estimación de mínimos cuadrados indirectos, es claro que los coeficientes en la forma reducida, las $\hat{\Pi}$, son funciones de las perturbaciones estocásticas, tales como $u_2$. Puesto que $\hat{Y}_{1t}$ depende de las $\hat{\Pi}$, es probable que esté correlacionada con $u_2$, que es un componente de $u_t^*$; como resultado, se espera que $\hat{Y}_{1t}$ esté correlacionada con $u_t^*$. Pero, como se mencionó, esta correlación desaparece a medida que el tamaño de la muestra tiende a infinito. La desventaja de todo esto es que en muestras pequeñas, el procedimiento de MC2E puede conducir a estimaciones sesgadas.

Como lo indica este procedimiento de dos etapas, la idea básica detrás de MC2E es "purificar" la variable explicativa estocástica $Y_1$ de la influencia de la perturbación estocástica $u_2$. Este objetivo se logra efectuando la regresión en forma reducida de $Y_1$ sobre todas las variables predeterminadas en el sistema (etapa 1), obteniendo las estimaciones $\hat{Y}_{1t}$ y reemplazando $Y_{1t}$ en la ecuación original por las $\hat{Y}_{1t}$ estimadas, para luego aplicar MCO a la ecuación así transformada (etapa 2). Los estimadores así obtenidos son consistentes, es decir, convergen hacia sus verdaderos valores a medida que el tamaño de la muestra aumenta indefinidamente.

Para ilustrar MC2E aún más, se modifica el modelo ingreso-oferta monetaria de la siguiente manera:

$$Y_{1t} = \beta_{10} + \beta_{12}Y_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \tag{20.4.7}$$

$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + \gamma_{23}X_{3t} + \gamma_{24}X_{4t} + u_{2t} \tag{20.4.8}$$

en donde, además de las variables ya definidas, $X_3 =$ el ingreso en el periodo anterior y $X_4 =$ la oferta monetaria en el periodo anterior. Tanto $X_3$ como $X_4$ son predeterminadas.

Puede verificarse fácilmente que ambas ecuaciones (20.4.7) y (20.4.8) están sobreidentificadas. Para aplicar MC2E se procede de la siguiente manera: en la etapa 1 se efectúa la regresión de las variables endógenas sobre todas las variables predeterminadas en el sistema. Así,

$$Y_{1t} = \hat{\Pi}_{10} + \hat{\Pi}_{11}X_{1t} + \hat{\Pi}_{12}X_{2t} + \hat{\Pi}_{13}X_{3t} + \hat{\Pi}_{14}X_{4t} + \hat{u}_{1t} \tag{20.4.9}$$

$$Y_{2t} = \hat{\Pi}_{20} + \hat{\Pi}_{21}X_{1t} + \hat{\Pi}_{22}X_{2t} + \hat{\Pi}_{23}X_{3t} + \hat{\Pi}_{24}X_{4t} + \hat{u}_{2t} \tag{20.4.10}$$

En la etapa 2 se remplazan $Y_1$ y $Y_2$ en las ecuaciones originales (estructurales) por sus valores estimados de las dos regresiones anteriores, y luego se efectúan las regresiones por MCO de la siguiente manera:

$$Y_{1t} = \beta_{10} + \beta_{12}\hat{Y}_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t}^* \tag{20.4.11}$$

$$Y_{2t} = \beta_{20} + \beta_{21}\hat{Y}_{1t} + \gamma_{23}X_{3t} + \gamma_{24}X_{4t} + u_{2t}^* \tag{20.4.12}$$

en donde $u_{1t}^* = u_{1t} + \beta_{12}\hat{u}_{2t}$ y $\hat{u}_{2t}^* = u_{2t} + \beta_{21}\hat{u}_{1t}$. Las estimaciones así obtenidas serán consistentes.

Observe las siguientes características de MC2E.

1. Puede aplicarse a una ecuación individual en el sistema sin tener en cuenta directamente ninguna otra ecuación o ecuaciones en el mismo. Por tanto, para resolver modelos econométricos que contienen un gran número de ecuaciones, MC2E ofrece un método económico. Por esta razón, el método ha sido extensamente utilizado en la práctica.
2. A diferencia de MCI, que proporciona múltiples estimaciones de los parámetros en las ecuaciones sobreidentificadas, MC2E proporciona solamente una estimación por parámetro.
3. Es fácil de aplicar porque todo lo que se necesita saber es el número total de variables exógenas o predeterminadas en el sistema sin conocer ninguna otra variable en el mismo.
4. Aunque está especialmente diseñado para manejar ecuaciones sobreidentificadas, el método también puede ser aplicado a ecuaciones exactamente identificadas. Pero entonces MCI y MC2E darán estimaciones idénticas. (¿Por qué?)
5. Si los valores de $R^2$ en las regresiones en forma reducida (es decir, regresiones de la primera etapa) son muy altos, por ejemplo, superan 0.8, las estimaciones clásicas por MCO y las de MC2E estarán muy cercanas. No obstante, este resultado no debe sorprender porque si el valor de $R^2$ en la primera etapa es muy alto, significa que los valores estimados de las variables endógenas están muy cercanos a sus valores observados y, por tanto, es menos probable que estas últimas estén correlacionadas con las perturbaciones estocásticas en las ecuaciones estructurales originales. (¿Por qué?)[15] Sin embargo, si los valores de $R^2$ en las regresiones de

[15] En caso extremo, si $R^2 = 1$ en la regresión de la primera etapa, la variable explicativa endógena en la ecuación original (sobreidentificada) será prácticamente no estocástica. (¿Por qué?)

la primera etapa son muy bajos, las estimaciones por MC2E prácticamente no tendrán significado porque se estarán remplazando las $Y$ originales en las regresiones de la segunda etapa por las $\hat{Y}$ estimadas en las regresiones de la primera etapa, las cuales representarán esencialmente las perturbaciones en las regresiones de esta etapa. En otras palabras, en este caso, las $\hat{Y}$ serán "variables representantes" muy deficientes de las $Y$ originales.

6. Observe que en los resultados de la regresión por MCI en la ecuación (20.3.15) no se determinaron los errores estándar de los coeficientes estimados (por las razones explicadas en la nota 10 de pie de página). Pero éstos pueden ser determinados para las estimaciones por MC2E puesto que los coeficientes estructurales son estimados directamente de las regresiones de la segunda etapa por MCO. Sin embargo, debe tenerse precaución en un aspecto. Los errores estándar estimados en las regresiones de la segunda etapa deben ser modificados porque, como se observó en la ecuación (20.4.6), el término de error $u_t^*$ es, en realidad, el término de error original, $u_{2t}$ más $\beta_{21}\hat{u}_t$. Por tanto, la varianza de $u_t^*$ no es exactamente igual a la varianza de $u_{2t}$ original. Sin embargo, la modificación requerida puede realizarse fácilmente mediante la fórmula dada en el apéndice 20A, sección 20A.2.
7. Al utilizar MC2E, tenga en mente las siguientes observaciones de Henri Theil:

> La justificación estadística de MC2E es del tipo de muestras grandes. Cuando no hay variables endógenas rezagadas,... los estimadores de los coeficientes por MC2E son consistentes si las variables exógenas son constantes en muestras repetidas y si las perturbaciones [que aparecen en las diversas ecuaciones de comportamiento o estructurales]... son independientes e idénticamente distribuidas con medias cero y varianzas finitas... Si estas dos condiciones se satisfacen, la distribución muestral de los estimadores de los coeficientes por MC2E se aproxima a la normal para muestras grandes...
>
> Cuando el sistema de ecuaciones contiene variables endógenas rezagadas, la consistencia y la normalidad de muestras grandes de los estimadores de los coeficientes por MC2E requieren una condición adicional,... que a medida que la muestra aumenta, el cuadrado de la media de los valores tomados por cada variable endógena rezagada converja en probabilidad hacia un límite positivo...
>
> Si [las perturbaciones que aparecen en las diversas ecuaciones estructurales] *no* están distribuidas independientemente, las variables endógenas rezagadas no son independientes del funcionamiento actual del sistema de ecuaciones..., lo que significa que estas variables no son realmente predeterminadas. Por consiguiente, si estas variables son consideradas como predeterminadas en el procedimiento de MC2E, los estimadores resultantes no son consistentes.[16]

## 20.5 MC2E: ejemplo numérico

Para ilustrar el método de MC2E, considere el modelo ingreso-oferta monetaria dado anteriormente en las ecuaciones (20.4.1) y (20.4.2). Como se mostró, la ecuación de la oferta monetaria está sobreidentificada. Para estimar los parámetros de esta ecuación, se acude al método de mínimos cuadrados en dos etapas. La información requerida para el análisis está dada en la tabla 20.2; esta tabla también contiene la información requerida para responder algunas de las preguntas hechas en los ejercicios.

### *Regresión de la etapa 1*

Primero se efectúa la regresión de la variable explicativa estocástica ingreso $Y_1$, representada por el PIB, sobre las variables predeterminadas, inversión privada $X_1$ y gastos del gobierno $X_2$, obteniendo los siguientes resultados:

$$\begin{aligned} \hat{Y}_{1t} &= 2\,689.848 \quad + \quad 1.8700X_{1t} + \quad 2.0343X_{2t} \\ \text{ee} &= \quad (67.9874) \qquad (0.1717) \qquad (0.1075) \\ t &= \quad (39.5639) \qquad (10.8938) \qquad (18.9295) \qquad R^2 = 0.9964 \end{aligned} \tag{20.5.1}$$

[16] Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp. 341-342.

**TABLA 20.2**
**PIB, M2, GASFED, BT6, Estados Unidos, 1970-2005**

Fuente: *Economic Report of the President*, 2007. Tablas B-2, B-69, B-84 y B-73.

| Observación | PIB ($Y_1$) | M2 ($Y_2$) | IDPB ($X_1$) | GASFED ($X_2$) | BT6 ($X_3$) |
|---|---|---|---|---|---|
| 1970 | 3 771.9 | 626.5 | 427.1 | 201.1 | 6.562 |
| 1971 | 3 898.6 | 710.3 | 475.7 | 220.0 | 4.511 |
| 1972 | 4 105.0 | 802.3 | 532.1 | 244.4 | 4.466 |
| 1973 | 4 341.5 | 855.5 | 594.4 | 261.7 | 7.178 |
| 1974 | 4 319.6 | 902.1 | 550.6 | 293.3 | 7.926 |
| 1975 | 4 311.2 | 1 016.2 | 453.1 | 346.2 | 6.122 |
| 1976 | 4 540.9 | 1 152.0 | 544.7 | 374.3 | 5.266 |
| 1977 | 4 750.5 | 1 270.3 | 627.0 | 407.5 | 5.510 |
| 1978 | 5 015.0 | 1 366.0 | 702.6 | 450.0 | 7.572 |
| 1979 | 5 173.4 | 1 473.7 | 725.0 | 497.5 | 10.017 |
| 1980 | 5 161.7 | 1 599.8 | 645.3 | 585.7 | 11.374 |
| 1981 | 5 291.7 | 1 755.4 | 704.9 | 672.7 | 13.776 |
| 1982 | 5 189.3 | 1 910.3 | 606.0 | 748.5 | 11.084 |
| 1983 | 5 423.8 | 2 126.5 | 662.5 | 815.4 | 8.75 |
| 1984 | 5 813.6 | 2 310.0 | 857.7 | 877.1 | 9.80 |
| 1985 | 6 053.7 | 2 495.7 | 849.7 | 948.2 | 7.66 |
| 1986 | 6 263.6 | 2 732.4 | 843.9 | 1 006.0 | 6.03 |
| 1987 | 6 475.1 | 2 831.4 | 870.0 | 1 041.6 | 6.05 |
| 1988 | 6 742.7 | 2 994.5 | 890.5 | 1 092.7 | 6.92 |
| 1989 | 6 981.4 | 3 158.5 | 926.2 | 1 167.5 | 8.04 |
| 1990 | 7 112.5 | 3 278.6 | 895.1 | 1 253.5 | 7.47 |
| 1991 | 7 100.5 | 3 379.1 | 822.2 | 1 315.0 | 5.49 |
| 1992 | 7 336.6 | 3 432.5 | 889.0 | 1 444.6 | 3.57 |
| 1993 | 7 532.7 | 3 484.0 | 968.3 | 1 496.0 | 3.14 |
| 1994 | 7 835.5 | 3 497.5 | 1 099.6 | 1 533.1 | 4.66 |
| 1995 | 8 031.7 | 3 640.4 | 1 134.0 | 1 603.5 | 5.59 |
| 1996 | 8 328.9 | 3 815.1 | 1 234.3 | 1 665.8 | 5.09 |
| 1997 | 8 703.5 | 4 031.6 | 1 387.7 | 1 708.9 | 5.18 |
| 1998 | 9 066.9 | 4 379.0 | 1 524.1 | 1 734.9 | 4.85 |
| 1999 | 9 470.3 | 4 641.1 | 1 642.6 | 1 787.6 | 4.76 |
| 2000 | 9 817.0 | 4 920.9 | 1 735.5 | 1 864.4 | 5.92 |
| 2001 | 9 890.7 | 5 430.3 | 1 598.4 | 1 969.5 | 3.39 |
| 2002 | 10 048.8 | 5 774.1 | 1 557.1 | 2 101.1 | 1.69 |
| 2003 | 10 301.0 | 6 062.0 | 1 613.1 | 2 252.1 | 1.06 |
| 2004 | 10 703.5 | 6 411.7 | 1 770.6 | 2 383.0 | 1.58 |
| 2005 | 11 048.6 | 6 669.4 | 1 866.3 | 2 555.9 | 3.40 |

*Notas:* $Y_1$ = PIB = producto interno bruto (miles de millones de dólares, ajustados al año 2000).
$Y_2$ = M2 = oferta de dinero M2 (miles de millones de dólares).
$X_1$ = IDPB = inversión doméstica privada bruta (miles de millones de dólares, ajustados al año 2000).
$X_2$ = GASFED = gasto del gobierno federal (miles de millones de dólares).
$X_3$ = BT6 = tasa de interés de los bonos del Tesoro a 6 meses (%).

*Regresión de la etapa 2*

Se estima ahora la función de oferta monetaria (20.4.2) remplazando la variable endógena $Y_1$ por $Y_1$ estimada de la ecuación (20.5.1) $(= \hat{Y}_1)$ . Los resultados son los siguientes:

$$
\begin{aligned}
\hat{Y}_{2t} &= -2\,440.180 \quad + \quad 0.7920\hat{Y}_{1t} \\
\text{ee} &= \quad (127.3720) \qquad (0.0178) \\
t &= \quad (-19.1579) \qquad (44.5246) \qquad R^2 = 0.9831
\end{aligned}
\tag{20.5.2}
$$

Ya se señaló que los errores estándar estimados dados en (20.5.2) deben ser corregidos en la forma sugerida en el apéndice 20.A, sección 20A.2. Una vez realizada esta corrección (la mayoría de los programas econométricos hacen esto rutinariamente), se obtienen los siguientes resultados:

$$\begin{aligned} \hat{Y}_{2t} &= -2\,440.180 \quad + \quad 0.7920\hat{Y}_{1t} \\ \text{ee} &= \quad (126.9598) \qquad (0.0212) \\ t &= \quad (-17.3149) \qquad (37.3057) \qquad R^2 = 0.9803 \end{aligned} \tag{20.5.3}$$

Como se menciona en el apéndice 20A, sección 20A.2, los errores estándar dados en (20.5.3) no difieren mucho de aquellos dados en (20.5.2), debido a que $R^2$ en la regresión de la primera etapa es muy alto.

### *Regresión por MCO*

Con fines comparativos, se presenta la regresión de las existencias de dinero sobre el ingreso, como se muestra en la ecuación (20.4.2), sin "depurar" la $Y_{1t}$ estocástica de la influencia del término de perturbación estocástica:

$$\begin{aligned} \hat{Y}_{2t} &= -2\,195.468 \quad + \quad 0.7911Y_{1t} \\ \text{ee} &= \quad (126.6460) \qquad (0.0211) \\ t &= \quad (-17.3354) \qquad (37.3812) \qquad R^2 = 0.9803 \end{aligned} \tag{20.5.4}$$

Al comparar los resultados "inapropiados" de MCO con la regresión de la etapa 2, se observa que las dos regresiones son virtualmente iguales. ¿Significa esto que el procedimiento de MC2E no tiene ningún valor? Definitivamente no. No debe sorprender que en la situación actual los dos resultados sean prácticamente idénticos porque, como se mencionó anteriormente, el valor del $R^2$ en la primera etapa es muy alto, igualando prácticamente $\hat{Y}_{1t}$ con $Y_{1t}$ observado. Por consiguiente, en este caso las regresiones por MCO y de la segunda etapa serán más o menos similares. Pero no hay garantía de que esto suceda en cada aplicación. Una implicación, entonces, es que en ecuaciones sobreidentificadas no debe aceptarse el procedimiento clásico de MCO sin verificar la(s) regresión(es) de la segunda etapa.

### *Simultaneidad entre el PIB y la oferta monetaria*

Para averiguar si el PIB ($Y_1$) y la oferta monetaria ($Y_2$) son mutuamente dependientes, se utiliza la prueba de simultaneidad de Hausman analizada en el capítulo 19.

Primero se efectúa la regresión del PIB sobre $X_1$ (gasto de inversión) y $X_2$ (gasto del gobierno), las variables exógenas en el sistema (es decir, se estima la regresión en la forma reducida). De esta regresión se obtiene el PIB estimado y los residuos $\hat{v}_t$, como lo indica la ecuación (19.4.7). Luego se efectúa la regresión de la oferta monetaria sobre el PIB estimado y sobre $v_t$ para obtener los siguientes resultados:

$$\begin{aligned} \hat{Y}_{2t} &= -2\,198.297 \quad + \quad 0.7915\hat{Y}_{1t} \; + \; 0.6984\hat{v}_t \\ \text{ee} &= \quad (129.0548) \qquad (0.0215) \qquad (0.2970) \\ t &= \quad (-17.0338) \qquad (36.70016) \qquad (2.3511) \end{aligned} \tag{20.5.5}$$

Puesto que el valor $t$ de $\hat{v}_t$ es estadísticamente significativo (el valor $p$ es 0.0263), no puede rechazarse la hipótesis de simultaneidad entre la oferta monetaria y el PIB, lo cual no debe sorprender. (*Nota*: en un sentido estricto, esta conclusión es válida sólo en muestras grandes, técnicamente, a medida que el tamaño de la muestra aumenta de manera indefinida.)

### *Pruebas de hipótesis*

Suponga que se desea probar la hipótesis de que el ingreso no tiene efecto sobre la demanda de dinero. ¿Se puede probar esta hipótesis con la prueba $t$ usual de la regresión estimada (20.5.2)? Sí, siempre y cuando la muestra sea grande y se corrijan los errores estándar, como se muestra en la ecuación (20.5.3); se puede utilizar la prueba $t$ para probar la significancia de un coeficiente individual y la prueba $F$ para probar la significancia conjunta de dos o más coeficientes, utilizando la fórmula (8.4.7).[17]

¿Qué sucede si el término de error en una ecuación estructural está autocorrelacionado y/o correlacionado con el término de error de otra ecuación estructural del sistema? Una respuesta completa a esta interrogante se sale del alcance de este libro y es mejor dejarla para las referencias (véase la referencia dada en la nota 7). Sin embargo, existen técnicas de estimación (tales como la técnica SURE de Zellner) para manejar estas complicaciones.

Para concluir el análisis de este ejemplo numérico, cabe añadir que los paquetes de software modernos, como STATA y *EViews,* realizan por lo regular los diferentes pasos que comprende la aplicación de MC2E. Sólo por motivos pedagógicos se demostraron los detalles de MC2E. Véase el ejercicio 20.15.

## 20.6 Ejemplos ilustrativos

En esta sección se consideran algunas aplicaciones de los métodos de ecuaciones simultáneas.

**EJEMPLO 20.1**
*Publicidad, concentración y márgenes de precios*

Para estudiar las interrelaciones entre la publicidad, la concentración (medida por la razón de concentración) y los márgenes precio-costo, Allyn D. Strickland y Leonard W. Weiss formularon el siguiente modelo de tres ecuaciones.[18]

*Función de intensidad de la publicidad*:

$$\text{Ad}/S = a_0 + a_1 M + a_2(\text{CD}/S) + a_3 C + a_4 C^2 + a_5\text{Gr} + a_6\text{Dur} \qquad \textbf{(20.6.1)}$$

*Función de concentración:*

$$C = b_0 + b_1(\text{Ad}/S) + b_2(\text{MES}/S) \qquad \textbf{(20.6.2)}$$

*Función del margen precio-costo:*

$$M = c_0 + c_1(K/S) + c_2\text{Gr} + c_3 C + c_4\text{GD} + c_5(\text{Ad}/S) + c_6(\text{MES}/S) \qquad \textbf{(20.6.3)}$$

en donde Ad = gasto en publicidad
S = valor de los embarques
C = razón de concentración de cuatro empresas
CD = demanda del consumidor
MES = escala mínima de eficiencia
M = margen precio/costo
Gr = tasa de crecimiento anual de la producción industrial
Dur = variable dicótoma para la industria de bienes durables
K = existencias de capital
GD = medida de la dispersión geográfica de la producción

[17] Pero hay que tener cuidado con lo siguiente: la SCR restringida y la no restringida en el numerador deben ser calculadas utilizando la $Y$ pronosticada (como en la etapa 2 de MC2E) y la SCR en el denominador está calculada utilizando valores observados y no pronosticados de las regresoras. Para una explicación sencilla a este respecto, véase T. Dudley Wallace y J. Lew Silver, *Econometrics: An Introduction*, Addison-Wesley, Reading, Mass., 1988, sec. 8.5.

[18] Véase su artículo "Advertising, Concentration, and Price-Cost Margins", *Journal of Political Economy*, vol. 84, núm. 5, 1976, pp. 1109-1121.

**EJEMPLO 20.1**
(*continuación*)

Según las condiciones de orden para la identificación, la ecuación (20.6.2) está sobreidentificada, en tanto que las ecuaciones (20.6.1) y (20.6.3) están exactamente identificadas.

La información para el análisis provino, en gran parte, del Census of Manufacturers de 1963 y cubrió 408 de las 417 industrias manufactureras de cuatro dígitos. Las tres ecuaciones fueron estimadas inicialmente por MCO, produciendo los resultados que aparecen en la tabla 20.3. Para corregir por el sesgo de las ecuaciones simultáneas, los autores reestimaron el modelo utilizando MC2E. Los resultados obtenidos se presentan en la tabla 20.4. Se deja al lector la comparación de los dos resultados.

**TABLA 20.3**
**Estimaciones por MCO de las tres ecuaciones (razones *t* entre paréntesis)**

| | Variable dependiente | | |
|---|---|---|---|
| | **Ad/*S*** **Ec. (20.6.1)** | ***C*** **Ec. (20.6.2)** | ***M*** **Ec. (20.6.3)** |
| Constante | −0.0314 (−7.45) | 0.2638 (25.93) | 0.1682 (17.15) |
| *C* | 0.0554 (3.56) | — | 0.0629 (2.89) |
| $C^2$ | −0.0568 (−3.38) | — | — |
| *M* | 0.1123 (9.84) | — | — |
| CD/*S* | 0.0257 (8.94) | — | — |
| Gr | 0.0387 (1.64) | | 0.2255 (2.61) |
| Dur | −0.0021 (−1.11) | — | — |
| Ad/*S* | — | 1.1613 (3.3) | 1.6536 (11.00) |
| MES/*S* | — | 4.1852 (18.99) | 0.0686 (0.54) |
| *K*/*S* | — | — | 0.1123 (8.03) |
| GD | — | — | −0.0003 (−2.90) |
| $R^2$ | 0.374 | 0.485 | 0.402 |
| gl | 401 | 405 | 401 |

**TABLA 20.4**
**Estimaciones de mínimos cuadrados en dos etapas de las tres ecuaciones (razones *t* entre paréntesis)**

| | Variable dependiente | | |
|---|---|---|---|
| | **Ad/*S*** **Ec. (20.6.1)** | ***C*** **Ec. (20.6.2)** | ***M*** **Ec. (20.6.3)** |
| Constante | −0.0245 (−3.86) | 0.2591 (21.30) | 0.1736 (14.66) |
| *C* | 0.0737 (2.84) | — | 0.0377 (0.93) |
| $C^2$ | −0.0643 (−2.64) | — | — |
| *M* | 0.0544 (2.01) | — | — |
| CD/*S* | 0.0269 (8.96) | — | — |
| Gr | 0.0539 (2.09) | — | 0.2336 (2.61) |
| Dur | −0.0018 (−0.93) | — | — |
| Ad/*S* | — | 1.5347 (2.42) | 1.6256 (5.52) |
| MES/*S* | — | 4.169 (18.84) | 0.1720 (0.92) |
| *K*/*S* | — | — | 0.1165 (7.30) |
| GD | — | — | −0.0003 (−2.79) |

**EJEMPLO 20.2**
*Modelo I de Klein*

En el ejemplo 18.6 se analizó, de manera breve, el modelo pionero de Klein. Inicialmente, el modelo fue estimado para el periodo 1920-1941. Los datos básicos se presentan en la tabla 20.5, y las estimaciones en la forma reducida por MCO y por MC2E están dadas en la tabla 20.6. Se deja al lector la interpretación de estos resultados.

(*continúa*)

**EJEMPLO 20.2** (*continuación*)

**TABLA 20.5** **Datos básicos para el modelo I de Klein**

| Año | C* | P | W | I | $K_{-1}$ | X | W′ | G | T |
|---|---|---|---|---|---|---|---|---|---|
| 1920 | 39.8 | 12.7 | 28.8 | 2.7 | 180.1 | 44.9 | 2.2 | 2.4 | 3.4 |
| 1921 | 41.9 | 12.4 | 25.5 | −0.2 | 182.8 | 45.6 | 2.7 | 3.9 | 7.7 |
| 1922 | 45.0 | 16.9 | 29.3 | 1.9 | 182.6 | 50.1 | 2.9 | 3.2 | 3.9 |
| 1923 | 49.2 | 18.4 | 34.1 | 5.2 | 184.5 | 57.2 | 2.9 | 2.8 | 4.7 |
| 1924 | 50.6 | 19.4 | 33.9 | 3.0 | 189.7 | 57.1 | 3.1 | 3.5 | 3.8 |
| 1925 | 52.6 | 20.1 | 35.4 | 5.1 | 192.7 | 61.0 | 3.2 | 3.3 | 5.5 |
| 1926 | 55.1 | 19.6 | 37.4 | 5.6 | 197.8 | 64.0 | 3.3 | 3.3 | 7.0 |
| 1927 | 56.2 | 19.8 | 37.9 | 4.2 | 203.4 | 64.4 | 3.6 | 4.0 | 6.7 |
| 1928 | 57.3 | 21.1 | 39.2 | 3.0 | 207.6 | 64.5 | 3.7 | 4.2 | 4.2 |
| 1929 | 57.8 | 21.7 | 41.3 | 5.1 | 210.6 | 67.0 | 4.0 | 4.1 | 4.0 |
| 1930 | 55.0 | 15.6 | 37.9 | 1.0 | 215.7 | 61.2 | 4.2 | 5.2 | 7.7 |
| 1931 | 50.9 | 11.4 | 34.5 | −3.4 | 216.7 | 53.4 | 4.8 | 5.9 | 7.5 |
| 1932 | 45.6 | 7.0 | 29.0 | −6.2 | 213.3 | 44.3 | 5.3 | 4.9 | 8.3 |
| 1933 | 46.5 | 11.2 | 28.5 | −5.1 | 207.1 | 45.1 | 5.6 | 3.7 | 5.4 |
| 1934 | 48.7 | 12.3 | 30.6 | −3.0 | 202.0 | 49.7 | 6.0 | 4.0 | 6.8 |
| 1935 | 51.3 | 14.0 | 33.2 | −1.3 | 199.0 | 54.4 | 6.1 | 4.4 | 7.2 |
| 1936 | 57.7 | 17.6 | 36.8 | 2.1 | 197.7 | 62.7 | 7.4 | 2.9 | 8.3 |
| 1937 | 58.7 | 17.3 | 41.0 | 2.0 | 199.8 | 65.0 | 6.7 | 4.3 | 6.7 |
| 1938 | 57.5 | 15.3 | 38.2 | −1.9 | 201.8 | 60.9 | 7.7 | 5.3 | 7.4 |
| 1939 | 61.6 | 19.0 | 41.6 | 1.3 | 199.9 | 69.5 | 7.8 | 6.6 | 8.9 |
| 1940 | 65.0 | 21.1 | 45.0 | 3.3 | 201.2 | 75.7 | 8.0 | 7.4 | 9.6 |
| 1941 | 69.7 | 23.5 | 53.3 | 4.9 | 204.5 | 88.4 | 8.5 | 13.8 | 11.6 |

* La interpretación de los encabezados de las columnas se describe en el ejemplo 18.6.

Fuente: Esta información se tomó de G. S. Maddala, *Econometrics*, McGraw-Hill, Nueva York, 1977, p. 238.

**TABLA 20.6*** **Estimaciones por MCO, forma reducida y MC2E del modelo I de Klein**

Fuente: G. S. Maddala, *Econometrics*, McGraw-Hill, Nueva York, 1977, p. 242.

**MCO:**

$$\begin{aligned}\hat{C} &= \underset{(1.203)}{16.237} + \underset{(0.091)}{0.193}P + \underset{(0.040)}{0.796}(W + W') + \underset{(0.090)}{0.089}P_{-1} \qquad \bar{R}^2 = 0.978 \quad \text{DW} = 1.367\\ \hat{I} &= \underset{(5.465)}{10.125} + \underset{(0.097)}{0.479}P + \underset{(0.100)}{0.333}P_{-1} - \underset{(0.026)}{0.112}K_{-1} \qquad \bar{R}^2 = 0.919 \quad \text{DW} = 1.810\\ \hat{W} &= \underset{(1.151)}{0.064} + \underset{(0.032)}{0.439}X + \underset{(0.037)}{0.146}X_{-1} + \underset{(0.031)}{0.130}t \qquad \bar{R}^2 = 0.985 \quad \text{DW} = 1.958\end{aligned}$$

**Forma reducida:**

$$\begin{aligned}\hat{P} &= \underset{(10.870)}{46.383} + \underset{(0.444)}{0.813}P_{-1} - \underset{(0.067)}{0.213}K_{-1} + \underset{(0.252)}{0.015}X_{-1} + \underset{(0.154)}{0.297}t - \underset{(0.385)}{0.926}T + \underset{(0.373)}{0.443}G\\ &\qquad \bar{R}^2 = 0.753 \quad \text{DW} = 1.854\\ \widehat{W + W'} &= \underset{(8.787)}{40.278} + \underset{(0.359)}{0.823}P_{-1} - \underset{(0.054)}{0.144}K_{-1} + \underset{(0.204)}{0.115}X_{-1} + \underset{(0.124)}{0.881}t - \underset{(0.311)}{0.567}T + \underset{(0.302)}{0.859}G\\ &\qquad \bar{R}^2 = 0.949 \quad \text{DW} = 2.395\\ \hat{X} &= \underset{(18.860)}{78.281} + \underset{(0.771)}{1.724}P_{-1} - \underset{(0.110)}{0.319}K_{-1} + \underset{(0.438)}{0.094}X_{-1} + \underset{(0.267)}{0.878}t - \underset{(0.669)}{0.565}T + \underset{(0.648)}{1.317}G\\ &\qquad \bar{R}^2 = 0.882 \quad \text{DW} = 2.049\end{aligned}$$

**MC2E:**

$$\begin{aligned}\hat{C} &= \underset{(1.464)}{16.543} + \underset{(0.130)}{0.019}P + \underset{(0.044)}{0.810}(W + W') + \underset{(0.118)}{0.214}P_{-1} \qquad \bar{R}^2 = 0.9726\\ \hat{I} &= \underset{(8.361)}{20.284} + \underset{(0.191)}{0.149}P + \underset{(0.180)}{0.616}P_{-1} - \underset{(0.040)}{0.157}K_{-1} \qquad \bar{R}^2 = 0.8643\\ \hat{W} &= \underset{(1.894)}{0.065} + \underset{(0.065)}{0.438}X + \underset{(0.070)}{0.146}X_{-1} + \underset{(0.053)}{0.130}t \qquad \bar{R}^2 = 0.9852\end{aligned}$$

*La interpretación de las variables se presenta en el ejemplo 18.6 (los errores estándar están entre paréntesis).

**EJEMPLO 20.3**
*Modelo de fijación de precios de activos de capital expresado como un sistema recursivo*

En una aplicación poco usual de los modelos recursivos de ecuaciones simultáneas, Cheng F. Lee y W. P. Lloyd[19] estimaron el siguiente modelo para la industria petrolera:

$$
\begin{aligned}
R_{1t} &= \alpha_1 + \gamma_1 M_t + u_{1t}\\
R_{2t} &= \alpha_2 + \beta_{21} R_{1t} + \gamma_2 M_t + u_{2t}\\
R_{3t} &= \alpha_3 + \beta_{31} R_{1t} + \beta_{32} R_{2t} + \gamma_3 M_t + u_{3t}\\
R_{4t} &= \alpha_4 + \beta_{41} R_{1t} + \beta_{42} R_{2t} + \beta_{43} R_{3t} + \gamma_4 M_t + u_{4t}\\
R_{5t} &= \alpha_5 + \beta_{51} R_{1t} + \beta_{52} R_{2t} + \beta_{53} R_{3t} + \beta_{54} R_{4t} + \gamma_5 M_t + u_{5t}\\
R_{6t} &= \alpha_6 + \beta_{61} R_{1t} + \beta_{62} R_{2t} + \beta_{63} R_{3t} + \beta_{64} R_{4t} + \beta_{65} R_{5t} + \gamma_6 M_t + u_{6t}\\
R_{7t} &= \alpha_7 + \beta_{71} R_{1t} + \beta_{72} R_{2t} + \beta_{73} R_{3t} + \beta_{74} R_{4t} + \beta_{75} R_{5t} + \beta_{76} R_{6t} + \gamma_7 M_t + u_{7t}
\end{aligned}
$$

donde $R_1$ = tasa de rendimiento sobre la acción 1 (= Imperial Oil)
$R_2$ = tasa de rendimiento sobre la acción 2 (= Sun Oil)
⋮
$R_7$ = tasa de rendimiento sobre la acción 7 (= Standard of Indiana)
$M_t$ = tasa de rendimiento sobre el índice de mercado
$u_{it}$ = perturbaciones ($i = 1, 2, \ldots, 7$)

Antes de presentar los resultados, la pregunta obvia es: ¿cómo se selecciona cuál es la acción 1, cuál es la acción 2, y así sucesivamente? Lee y Lloyd responden esta pregunta en forma empírica. Efectúan la regresión de la tasa de rendimiento de la acción *i* sobre las tasas de rendimiento de las seis acciones restantes y observan la $R^2$ resultante. Así, obtienen siete regresiones de este tipo. Entonces ordenan los valores de la $R^2$ estimados, de menor a mayor. La acción que tiene la menor $R^2$ se selecciona como la acción 1 y la que tiene la $R^2$ mayor se selecciona como la 7. La idea detrás de esto es intuitivamente simple. Si la $R^2$ de la tasa de rendimiento de, por ejemplo, Imperial Oil es la más baja con respecto a las seis acciones restantes, esto indica que esta acción es la menos afectada por los movimientos en los rendimientos de las demás acciones. Por consiguiente, el ordenamiento causal, de existir, va de esta acción hacia las otras, y no hay realimentación procedente de las otras acciones.

Aunque puede objetarse este enfoque puramente experimental del ordenamiento causal, en la tabla 20.7 se presentan sus resultados empíricos.

En el ejercicio 5.5 se introdujo la *línea característica* de la teoría moderna de la inversión, que es simplemente la regresión de la tasa de rendimiento de la acción *i* sobre la tasa de rendimiento del mercado. El coeficiente de la pendiente, conocido como *coeficiente beta,* es una medida de la volatilidad del rendimiento de la acción. Los resultados de la regresión de Lee-Lloyd proponen que hay relaciones intraindustriales significativas entre los rendimientos de las acciones, además de la influencia común del mercado representadas por la cartera del mercado. Así, el rendimiento de Standard de Indiana depende no solamente de la tasa de rendimiento del mercado sino, además, de las tasas de rendimiento de Shell Oil, Phillips Petroleum y Union Oil. Para plantear la idea en forma diferente, el movimiento de la tasa de rendimiento de Standard de Indiana puede explicarse mejor si, además de la tasa de rendimiento del mercado, se consideran las tasas de rendimiento obtenidas por Shell Oil, Phillips Petroleum y Union Oil.

(*continúa*)

[19] "The Capital Asset Pricing Model Expressed as a Recursive System: An Empirical Investigation", *Journal of Financial and Quantitative Analysis*, junio de 1976, pp. 237-249.

**EJEMPLO 20.3**
(*continuación*)

**TABLA 20.7 Estimaciones del sistema recursivo para la industria petrolera**

| | Forma lineal Variable dependientes | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Standard of Indiana** | **Shell Oil** | **Phillips Petroleum** | **Union Oil** | **Standard of Ohio** | **Sun Oil** | **Imperial Oil** |
| Standard of Indiana | | | | | | | |
| Shell Oil | 0.2100* | | | | | | |
| | (2.859) | | | | | | |
| Phillips Petroleum | 0.2293* | 0.0791 | | | | | |
| | (2.176) | (1.065) | | | | | |
| Union Oil | 0.1754* | 0.2171* | 0.2225* | | | | |
| | (2.472) | (3.177) | (2.337) | | | | |
| Standard of Ohio | −0.0794 | 0.0147 | 0.4248* | 0.1468* | | | |
| | (−1.294) | (0.235) | (5.501) | (1.735) | | | |
| Sun Oil | 0.1249 | 0.1710* | 0.0472 | 0.1339 | 0.0499 | | |
| | (1.343) | (1.843) | (0.355) | (0.908) | (0.271) | | |
| Imperial Oil | −0.1077 | 0.0526 | 0.0354 | 0.1580 | −0.2541* | 0.0828 | |
| | (−1.412) | (0.6804) | (0.319) | (1.290) | (−1.691) | (0.971) | |
| Constante | 0.0868 | −0.0384 | −0.0127 | −0.2034 | 0.3009 | 0.2013 | 0.3710* |
| | (0.681) | (1.296) | (−0.068) | (0.986) | (1.204) | (1.399) | (2.161) |
| Índice de mercado | 0.3681* | 0.4997* | 0.2884 | 0.7609* | 0.9089* | 0.7161* | 0.6432* |
| | (2.165) | (3.039) | (1.232) | (3.069) | (3.094) | (4.783) | (3.774) |
| $R^2$ | 0.5020 | 0.4658 | 0.4106 | 0.2532 | 0.0985 | 0.2404 | 0.1247 |
| Durbin-Watson | 2.1083 | 2.4714 | 2.2306 | 2.3468 | 2.2181 | 2.3109 | 1.9592 |

*Denota significancia al nivel del 0.10 o mejor para la prueba de dos colas.
*Nota:* Los valores *t* aparecen entre paréntesis debajo de los coeficientes.
Fuente: Cheng F. Lee y W.P. Lloyd, *op. cit.*, tabla 3b.

**EJEMPLO 20.4**
*Forma revisada del modelo de San Luis*[20]

El conocido y controvertido modelo de San Luis, desarrollado originalmente a finales del decenio de los años sesenta, ha sido revisado con alguna frecuencia. Una de tales revisiones se presenta en la tabla 20.8 y los resultados empíricos basados en este modelo revisado están dados en la tabla 20.9. (*Nota:* Un punto sobre una variable representa la tasa de crecimiento de dicha variable.) El modelo consiste básicamente en las ecuaciones (1), (2), (4) y (5) de la tabla 20.8; las otras ecuaciones representan las definiciones. La ecuación (1) fue estimada por MCO. Las ecuaciones (1), (2) y (4) fueron estimadas utilizando el método de rezagos distribuidos de Almon con restricciones (de punto final) sobre los coeficientes. En donde es pertinente, las ecuaciones fueron corregidas por correlación serial de primer orden ($\rho_1$) y/o correlación serial de segundo orden ($\rho_2$).

Al examinar los resultados se observa que la tasa de crecimiento de la oferta monetaria es el determinante principal de la tasa de crecimiento del PIB (nominal) y no la tasa de crecimiento de los gastos que generan un alto nivel de empleo. La suma de los *M* coeficientes es 1.06, lo cual indica que un incremento (sostenido) de 1% en la oferta monetaria conduce, en promedio, a cerca de 1.06% de incremento en el PIB nominal. Por otra parte, la suma de los *E* coeficientes, cercana a 0.05, indica que un cambio en el gasto que propicia niveles altos de empleo por parte del gobierno tiene poco impacto sobre la tasa de crecimiento del PIB nominal. Se deja al lector interpretar los resultados de las demás regresiones que aparecen en la tabla 20.9.

[20] Federal Reserve Bank of St. Louis, *Review*, mayo de 1982, p. 14.

**EJEMPLO 20.4**
(*continuación*)

## TABLA 20.8 El modelo de San Luis

(1) $$\dot{Y}_1 = C1 + \sum_{i=0}^{4} CM_i(\dot{M}_{t-i}) + \sum_{i=0}^{4} CE(\dot{E}_{t-i}) + \varepsilon 1_t$$

(2) $$\dot{P}_t = C2 + \sum_{i=1}^{4} CPE_i(\dot{PE}_{t-i}) + \sum_{i=0}^{5} CD_i(\dot{X}_{t-i} - \dot{XF}^{*}_{t-i1})$$
$$+ CPA(\dot{PA}_t) + CDUM1(DUM1) + CDUM2(DUM2) + \varepsilon 2_t$$

(3) $$\dot{PA}_t = \sum_{i=1}^{21} CPRL_i(\dot{P}_{t-i})$$

(4) $$RL_t = C3 + \sum_{i=0}^{20} CPRL_i(\dot{P}_{t-i}) + \varepsilon 3_t$$

(5) $$U_t - UF_t = CG(GAP_t) + CG1(GAP_{t-1}) + \varepsilon 4_t$$

(6) $$Y_t = (P_t/100)(X_t)$$

(7) $$\dot{Y}_t = [(Y_t/Y_{t-i})^4 - 1]100$$

(8) $$\dot{X}_t = [(X_t/X_{t-i})^4 - 1]100$$

(9) $$\dot{P}_t = [(P_t/P_{t-i})^4 - 1]100$$

(10) $$GAP_t = [(XF_t/X_t)/XF_t]100$$

(11) $$\dot{XF}^{*}_t = [(XF_t/X_{t-1})^4 - 1]100$$

$Y$ = PIB nominal
$M$ = existencias de dinero (M1)
$E$ = gastos que arrojan un alto nivel de empleo
$P$ = deflactor del PIB (1972 = 100)
PE = precio relativo de la energía
$X$ = producción en dólares de 1972
XF = producción potencial (Rasche/Tatom)

RL = tasa de los bonos empresariales
$U$ = tasa de desempleo
UF = tasa de desempleo en pleno empleo
DUM1 = variable dicótoma de control (1971-III a 1973-I = 1; 0 en los demás casos)
DUM2 = variable dicótoma de control posterior (1973-II a 1975-I = 1; 0 en los demás casos)

Fuente: Federal Reserve Bank of St. Louis, *Review*, mayo de 1982, p. 14.

## TABLA 20.9

**Estimación dentro de la muestra: 1960-I a 1980-IV (el valor absoluto del estadístico *t* está entre paréntesis)**

Fuente: Federal Reserve Bank of St. Louis, mayo de 1982, p. 14.

(1)
$$\hat{\dot{Y}}_t = \underset{(2.15)}{2.44} + \underset{(3.38)}{0.40}\dot{M}_t + \underset{(5.06)}{0.39}\dot{M}_{t-1} + \underset{(2.18)}{0.22}\dot{M}_{t-2} + \underset{(0.82)}{0.06}\dot{M}_{t-3} - \underset{(0.11)}{0.01}\dot{M}_{t-4}$$
$$+ \underset{(1.46)}{0.06}\dot{E}_t + \underset{(0.63)}{0.02}\dot{E}_{t-1} - \underset{(0.57)}{0.02}\dot{E}_{t-2} - \underset{(0.52)}{0.02}\dot{E}_{t-3} + \underset{(0.34)}{0.01}\dot{E}_{t-4}$$
$$R^2 = 0.39 \qquad ee = 3.50 \qquad DW = 2.02$$

(2)
$$\hat{\dot{P}}_t = \underset{(2.53)}{0.96} + \underset{(0.75)}{0.01}\dot{PE}_{t-1} + \underset{(1.96)}{0.04}\dot{PE}_{t-2} - \underset{(0.73)}{0.01}\dot{PE}_{t-3} + \underset{(1.38)}{0.02}\dot{PE}_{t-4}$$
$$- \underset{(0.18)}{0.00}(\dot{X}_t - \dot{XF}^{*}_t) + \underset{(1.43)}{0.01}(\dot{X}_{t-1} - \dot{XF}^{*}_{t-1}) + \underset{(4.63)}{0.02}(\dot{X}_{t-2} - \dot{XF}^{*}_{t-2})$$
$$+ \underset{(3.00)}{0(02}(\dot{X}_{t-3} - \dot{XF}^{*}_{t-3}) + \underset{(2.42)}{0.02}(\dot{X}_{t-4} - \dot{XF}^{*}_{t-4} + \underset{(2.16)}{0(01}(\dot{X}_{t-5} - \dot{XF}^{*}_{t-5})$$
$$+ \underset{(10.49)}{1(03}(\dot{PA}_t) - \underset{(1.02)}{0.61}(DUM1_t) + \underset{(2.71)}{1.65}(DUM2_t)$$
$$R^2 = 0.80 \qquad ee = 1.28 \qquad DW = 1.97 \qquad \hat{\rho} = 0.12$$

(4)
$$\widehat{RL}_t = \underset{(3.12)}{2.97} + \underset{(5.22)}{0.96}\sum_{i=0}^{20} \dot{P}_{t-i}$$
$$R^2 = 0.32 \qquad ee = 0.33 \qquad DW = 1.76 \qquad \hat{\rho} = 0.94$$

(5)
$$\widehat{U_t - UF}_t = \underset{(11.89)}{0.28}(GAP_t) + \underset{(6.31)}{0.14}(GAP_{t-1})$$
$$R^2 = 0.63 \qquad ee = 0.17 \qquad DW = 1.95 \qquad \hat{\rho}_1 = 1.43 \qquad \hat{\rho}_2 = 0.52$$

## Resumen y conclusiones

1. Si en un modelo de ecuaciones simultáneas una ecuación está identificada (en forma exacta o sobreidentificada), se dispone de diversos métodos para estimarla.
2. Estos métodos se clasifican en dos categorías generales: *métodos uniecuacionales* y *métodos de sistemas*.
3. Por razones de economía, errores de especificación, etc., los métodos uniecuacionales son los más comunes. Una característica única de estos métodos es que es posible estimar aisladamente una ecuación que forma parte de un modelo multiecuacional sin preocuparse mucho de las otras ecuaciones del sistema. (*Nota*: Para fines de identificación, sin embargo, las demás ecuaciones en el sistema sí cuentan.)
4. Tres métodos uniecuacionales comúnmente utilizados son: **MCO, MCI** y **MC2E**.
5. Aunque el de MCO en general es inapropiado en el contexto de los modelos de ecuaciones simultáneas, puede ser aplicado a los **modelos recursivos** en donde hay una relación causa y efecto definida pero unidireccional entre las variables endógenas.
6. El método de MCI es apropiado para ecuaciones precisas o exactamente identificadas. Mediante este método, se aplica MCO a la ecuación en la forma reducida, y es a partir de los coeficientes de dicha forma que se estiman los coeficientes estructurales originales.
7. El método de MC2E está diseñado en especial para ecuaciones sobreidentificadas, aunque también puede aplicarse a ecuaciones exactamente identificadas. Pero entonces los resultados de MC2E y MCI son idénticos. La idea básica detrás de MC2E es reemplazar la variable explicativa endógena (estocástica) por una combinación lineal de variables predeterminadas en el modelo y utilizar esta combinación como variable explicativa en lugar de la variable endógena original. El método MC2E se parece entonces al **método de estimación de variables instrumentales**, en el cual la combinación lineal de las variables predeterminadas sirve como instrumento o variable representante para la regresora endógena.
8. Una característica importante sobre MCI y MC2E es que las estimaciones obtenidas son consistentes; es decir, a medida que el tamaño de la muestra aumenta indefinidamente, las estimaciones convergen hacia sus verdaderos valores poblacionales. Las estimaciones pueden no satisfacer las propiedades de muestra pequeña tales como el insesgamiento y la varianza mínima. Por consiguiente, los resultados obtenidos mediante la aplicación de estos métodos a muestras pequeñas, así como las inferencias obtenidas de ellos deben ser interpretados con la debida precaución.

## EJERCICIOS

### Preguntas

20.1 Determine si cada una de las siguientes afirmaciones es verdadera o falsa:

*a*) El método de MCO no es aplicable para estimar una ecuación estructural en un modelo de ecuaciones simultáneas.

*b*) En caso de que una ecuación no sea identificada, MC2E no es aplicable.

*c*) El problema de la simultaneidad no surge en un modelo recursivo de ecuaciones simultáneas.

*d*) Los problemas de simultaneidad y de exogeneidad significan lo mismo.

*e*) El método de MC2E y otros métodos de estimación de ecuaciones estructurales tienen propiedades estadísticas deseables solamente en muestras grandes.

*f*) En los modelos de ecuaciones simultáneas no existe un concepto similar al de $R^2$.

**g*) El método de MC2E y otros métodos de estimación de ecuaciones estructurales no son aplicables si los errores de la ecuación están autocorrelacionados y/o están correlacionados entre ecuaciones.

*h*) Si una ecuación está exactamente identificada, MCI y MC2E dan resultados idénticos.

* Opcional.

20.2. ¿Por qué no es necesario aplicar el método de mínimos cuadrados en dos etapas a ecuaciones exactamente identificadas?

20.3. Considere el siguiente modelo keynesiano modificado de determinación del ingreso:

$$C_t = \beta_{10} + \beta_{11} Y_t + u_{1t}$$
$$I_t = \beta_{20} + \beta_{21} Y_t + \beta_{22} Y_{t-1} + u_{2t}$$
$$Y_t = C_t + I_t + G_t$$

en donde $C =$ gasto de consumo
$I =$ gasto de inversión
$Y =$ ingreso
$G =$ gasto del gobierno
$G_t$ y $Y_{t-1}$ se suponen predeterminados.

*a*) Obtenga las ecuaciones en forma reducida y determine cuáles de las ecuaciones anteriores están identificadas (en forma exacta o sobreidentificadas).

*b*) ¿Cuál método puede utilizarse para estimar los parámetros de la ecuación sobreidentificada y de la ecuación exactamente identificada? Justifique la respuesta.

20.4. Considere los siguientes resultados:*

$$\text{MCO:}\quad \widehat{\dot{W}_t} = 0.276 + 0.258\dot{P}_t + 0.046\dot{P}_{t-1} + 4.959V_t \qquad R^2 = 0.924$$
$$\text{MCO:}\quad \widehat{\dot{P}_t} = 2.693 + 0.232\dot{W}_t - 0.544\dot{X}_t + 0.247\dot{M}_t + 0.064\dot{M}_{t-1} \qquad R^2 = 0.982$$
$$\text{MC2E:}\quad \widehat{\dot{W}_t} = 0.272 + 0.257\dot{P}_t + 0.046\dot{P}_{t-1} + 4.966V_t \qquad R^2 = 0.920$$
$$\text{MC2E:}\quad \widehat{\dot{P}_t} = 2.686 + 0.233\dot{W}_t - 0.544\dot{X}_t + 0.246\dot{M}_t + 0.046\dot{M}_{t-1} \qquad R^2 = 0.981$$

en donde $\dot{W}_t$, $\dot{P}_t$, $\dot{M}_t$ y $\dot{X}_t$ son los cambios porcentuales en las ganancias, los precios, los precios de importación y la productividad laboral (todos los cambios porcentuales se calculan con base en el año anterior), respectivamente, y donde $V_t$ representa las vacantes de empleo sin ocupar (porcentaje del número total de empleados).

"Puesto que los resultados de MCO y MC2E son prácticamente idénticos, MC2E no tiene ningún valor." Comente.

†20.5. Suponga que la producción está caracterizada por la función de producción Cobb-Douglas

$$Q_i = AK_i^{\alpha} L_i^{\beta}$$

en donde $Q =$ producción
$K =$ insumo capital
$L =$ insumo trabajo
$A$, $\alpha$ y $\beta =$ parámetros
$i = i$-ésima empresa

Dado el precio de la producción final $P$, el precio del trabajo $W$ y el precio del capital $R$, y suponiendo maximización de utilidades, se obtiene el siguiente modelo empírico de producción:

*Función de producción:*

$$\ln Q_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln u_{1i} \tag{1}$$

---

* Fuente: *Prices and Earnings in 1951-1969: An Econometric Assessment*, Department of Employment, Reino Unido, Her Majesty's Stationery Office, Londres, 1971, p. 30.

† Opcional.

*Función del producto marginal del trabajo*:

$$\ln Q_i = -\ln \beta + \ln L_i + \ln \frac{W}{P} + \ln u_{2i} \tag{2}$$

*Función del producto marginal del capital*:

$$\ln Q_i = -\ln \alpha + \ln K_i + \ln \frac{R}{P} + \ln u_{3i} \tag{3}$$

en donde $u_1$, $u_2$ y $u_3$ son las perturbaciones estocásticas.

En el modelo anterior hay tres ecuaciones con tres variables endógenas $Q$, $L$ y $K$. $P$, $R$ y $W$ son exógenas.

*a*) ¿Cuáles problemas se encuentran al estimar el modelo si $\alpha + \beta = 1$, es decir, cuando hay rendimientos constantes a escala?

*b*) Aun si $\alpha + \beta \neq 1$, ¿se pueden estimar las ecuaciones? Responda considerando la identificación del sistema.

*c*) Si el sistema no está identificado, ¿qué puede hacerse para convertirlo en identificable?

*Nota*: Las ecuaciones (2) y (3) se obtienen al diferenciar $Q$ con respecto al trabajo y al capital, respectivamente, igualándolos a $W/P$ y $R/P$, transformando las expresiones resultantes en logaritmos y agregando (el logaritmo de) los términos de perturbación.

20.6. Considere el siguiente modelo de demanda y oferta de dinero:

*Demanda de dinero:* $M_t^d = \beta_0 + \beta_1 Y_1 + \beta_2 R_t + \beta_3 P_t + u_{1t}$

*Oferta de dinero:* $M_t^s = \alpha_0 + \alpha_1 Y_t + u_{2t}$

en donde $M =$ dinero
$Y =$ ingreso
$R =$ tasa de interés
$P =$ precio

Suponga que $R$ y $P$ están predeterminados.

*a*) ¿Está identificada la función de demanda?

*b*) ¿Está identificada la función de oferta?

*c*) ¿Cuál método se utilizaría para estimar los parámetros de la(s) ecuación(es) identificada(s)? ¿Por qué?

*d*) Suponga que se modifica la función de oferta agregando las variables explicativas $Y_{t-1}$ y $M_{t-1}$. ¿Qué sucede con el problema de la identificación? ¿Se utilizaría aún el método que utilizó en (*c*)? ¿Por qué sí o por qué no?

20.7. Remítase al ejercicio 18.10. Para el sistema de dos ecuaciones se obtuvieron las ecuaciones en forma reducida y se estimaron sus parámetros. Estime la regresión por mínimos cuadrados indirectos del consumo sobre el ingreso y compare los resultados con la regresión por MCO.

## Ejercicios empíricos

20.8. Considere el siguiente modelo:

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$$

en donde $M_t$ (oferta monetaria) es exógena, $R_t$ es la tasa de interés y $Y_t$ es el PIB.

*a*) ¿Cómo se justificaría el modelo?

*b*) ¿Están identificadas las ecuaciones?

*c*) Con la información dada en la tabla 20.2, estime los parámetros de las ecuaciones identificadas. Justifique el (los) método(s) que se utiliza(n).

20.9. Suponga que en el ejercicio 20.8 se cambia el modelo de la siguiente manera:

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + \beta_3 Y_{t-1} + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$$

*a*) Averigüe si el sistema está identificado.

*b*) Con la información dada en la tabla 20.2, estime los parámetros de la(s) ecuación(es) identificada(s).

20.10. Considere el siguiente modelo:

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$$

en donde las variables están definidas como en el ejercicio 20.8. Al considerar $I$ (inversión doméstica) y $M$ exógenamente, determine la identificación del sistema. Utilizando la información de la tabla 20.2, estime los parámetros de la(s) ecuación(es) identificada(s).

20.11. Suponga que se cambia el modelo del ejercicio 20.10 de la siguiente manera:

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$$
$$I_t = \gamma_0 + \gamma_1 R_t + u_{3t}$$

Suponga que $M$ está determinado exógenamente.

*a*) Determine cuáles ecuaciones están identificadas.

*b*) Estime los parámetros de la(s) ecuación(es) identificada(s) utilizando la información de la tabla 20.2. Justifique el (los) método(s).

20.12. Verifique los errores estándar presentados en (20.5.3).

20.13. Remítase al modelo de demanda y oferta dado en las ecuaciones (20.3.1) y (20.3.2). Suponga que la función de oferta se altera de la siguiente manera:

$$Q_t = \beta_0 + \beta_1 P_{t-1} + u_{2t}$$

en donde $P_{t-1}$ es el precio predominante en el periodo anterior.

*a*) Si $X$ (gasto) y $P_{t-1}$ están predeterminadas, ¿existe un problema de simultaneidad?

*b*) Si existe, ¿están determinadas cada una de las funciones de demanda y de oferta? Si lo están, obtenga las ecuaciones en forma reducida y estímelas con base en la información dada en la tabla 20.1.

*c*) ¿Pueden derivarse los coeficientes estructurales a partir de los coeficientes en la forma reducida? Muestre los cálculos necesarios.

20.14. *Ejercicio en clase*: Considere el siguiente modelo macroeconómico simple para la economía estadounidense durante el periodo 1960-1999.*

*Función consumo privado:*

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t} \qquad \alpha_1 > 0,\ 0 < \alpha_2 < 1$$

*Función inversión privada bruta:*

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 R_t + \beta_3 I_{t-1} + u_{2t} \qquad \beta_1 > 0,\ \beta_2 < 0,\ 0 < \beta_3 < 1$$

*Función de demanda de dinero:*

$$R_t = \lambda_0 + \lambda_1 Y_t + \lambda_2 M_{t-1} + \lambda_3 P_t + \lambda_4 R_{t-1} + u_{3t}$$
$$\lambda_1 > 0,\ \lambda_2 < 0,\ \lambda_3 > 0,\ 0 < \lambda_4 < 1$$

* Adaptado de H.R. Seddighi, K.A. Lawler y A.V. Katos, *Econometrics: A Practical Approach*, Routledge, Nueva York, 2000, p. 204.

*Identidad de ingreso:*

$$Y_t = C_t + I_t + G_t$$

en donde $C$ = consumo privado real, $I$ = inversión privada bruta real, $G$ = gasto gubernamental real, $Y$ = PIB real, $M$ = oferta de dinero M2 a precios actuales, $R$ = tasa de interés a largo plazo (%) y $P$ = índice de precios al consumidor. Las variables endógenas son $C$, $I$, $R$ y $Y$. Las variables predeterminadas son: $C_{t-1}$, $I_{t-1}$, $M_{t-1}$, $P_t$, $R_{t-1}$ y $G_t$ más el término del intercepto. Las $u$ son los términos de error.

*a*) Al utilizar la condición de orden para la identificación, determine cuál de las cuatro ecuaciones es exactamente identificada o sobreidentificada.

*b*) ¿Qué método(s) se utiliza(n) para calcular las ecuaciones identificadas?

*c*) Obtenga datos apropiados de fuentes gubernamentales y/o privadas, estime el modelo y comente los resultados.

20.15. En este ejercicio se examinarán los datos de 534 trabajadores, obtenidos de la *Current Population Survey* (CPS) correspondiente a 1985. Los datos se presentan en la tabla 20.10 del sitio web* del libro de texto. Las variables de esta tabla se definen como sigue:

$W$ = salarios \$, por hora; occup = ocupación; sector = 1 para industria manufacturera, 2 para construcción, 0 para otras; union = 1 si es trabajador sindicalizado, 0 en caso contrario; educ = años de escolaridad; exper = experiencia laboral en años; age = edad en años; sex = 1 para mujeres; marital status = 1 si es casado; race = 1 para otros, 2 para hispanos, 3 para blancos; region = 1 si vive en el sur de Estados Unidos.

Considere el siguiente modelo sencillo de determinación de salarios:

$$\ln W = \beta_1 + \beta_2 \text{Educ} + \beta_3 \text{Exper} + \beta_4 \text{Exper}^2 + u_i \qquad \textbf{(1)}$$

*a*) Suponga que la escolaridad, al igual que los salarios, es endógena. ¿Cómo se puede averiguar si en la ecuación (1) la escolaridad es, en efecto, endógena? Use los datos presentados en la tabla para el análisis.

*b*) ¿La prueba de Hausman sustenta el análisis realizado en (*a*)? Proporcione una explicación exhaustiva.

20.16. *Ejercicio en clase:* Considere el siguiente modelo de demanda y oferta para préstamos de bancos comerciales a empresas:

$$\textit{Demanda: } Q_t^d = \alpha_1 + \alpha_2 R_t + \alpha_2 \text{RD}_t + \alpha_4 \text{IPI}_t + u_{1t}$$

$$\textit{Oferta: } Q_t^s = \beta_1 + \beta_2 R_t + \beta_3 \text{RS}_t + \beta_4 \text{TBD}_t + u_{2t}$$

en donde $Q$ = total de préstamos de bancos comerciales (miles de millones de dólares); $R$ = tasa preferencial promedio; $RS$ = tasa de interés de los pagarés del Tesoro a 3 meses; $RD$ = calificación AAA de los bonos corporativos; $IPI$ = índice de producción industrial; y $TBD$ = total de depósitos bancarios.

*a*) Recopile datos sobre estas variables para el periodo 1980-2007 de varias fuentes, como www.economagic.com, el sitio Web del Banco de la Reserva Federal de San Luis, o cualquier otra fuente.

*b*) ¿Están identificadas las funciones de demanda y oferta? Indique cuáles variables son endógenas y cuáles son exógenas.

*c*) ¿Cómo podría estimar las funciones de demanda y oferta indicadas con anterioridad? Realice los cálculos necesarios.

*d*) ¿Por qué se incluyen $R$ y $RS$ en el modelo? ¿Qué papel desempeña $IPI$ en el modelo?

* Los datos se encuentran publicados en la Web, en http://lib.stat.cmu.edu/datasets/cps_85_wages.

# Apéndice 20A

## 20A.1 Sesgo en los estimadores de mínimos cuadrados indirectos

Para mostrar que los estimadores de MCI, aunque consistentes, están sesgados, se utiliza el modelo de demanda y oferta dado en las ecuaciones (20.3.1) y (20.3.2). De (20.3.10) se obtiene

$$\hat{\beta}_1 = \frac{\hat{\Pi}_3}{\hat{\Pi}_1}$$

Ahora

$$\hat{\Pi}_3 = \frac{\sum q_t x_t}{\sum x_t^2} \qquad \text{de la ecuación (20.3.7)}$$

y

$$\hat{\Pi}_1 = \frac{\sum p_t x_t}{\sum x_t^2} \qquad \text{de la ecuación(20.3.5)}$$

Por consiguiente, al sustituir, se obtiene

$$\hat{\beta}_1 = \frac{\sum q_t x_t}{\sum p_t x_t} \tag{1}$$

Al utilizar (20.3.3) y (20.3.4), se obtiene

$$p_t = \Pi_1 x_t + (w_t - \bar{w}) \tag{2}$$

$$q_t = \Pi_3 x_t + (v_t - \bar{v}) \tag{3}$$

en donde $\bar{w}$ y $\bar{v}$ son los valores de las medias de $w_t$ y $v_t$, respectivamente.

Al sustituir (2) y (3) en (1), se obtiene

$$\begin{aligned}\hat{\beta}_1 &= \frac{\Pi_3 \sum x_t^2 + \sum (v_t - \bar{v}) x_t}{\Pi_1 \sum x_t^2 + \sum (w_t - \bar{w}) x_t} \\ &= \frac{\Pi_3 + \sum (v_t - \bar{v}) x_t) \sum x_t^2}{\Pi_1 + \sum (w_t - \bar{w}) x_t) \sum x_t^2}\end{aligned} \tag{4}$$

Puesto que el operador $E$ de valor esperado es un operador lineal, no se puede tomar la esperanza de (4), aunque está claro que generalmente $\hat{\beta}_1 \neq (\Pi_3/\Pi_1)$. (¿Por qué?)

Pero, a medida que el tamaño de la muestra tiende a infinito, se puede obtener

$$\text{plím}\,(\hat{\beta}_1) = \frac{\text{plím}\,\Pi_3 + \text{plím}\,\sum (v_t - \bar{v}) x_t) \sum x_t^2}{\text{plím}\,\Pi_1 + \text{plím}\,\sum (w_t - \bar{w}) x_t) \sum x_t^2} \tag{5}$$

en donde se hace uso de las propiedades de plím, a saber:

$$\text{plím}\,(A + B) = \text{plím}\,A + \text{plím}\,B \text{ y plím}\left(\frac{A}{B}\right) = \frac{\text{plím}\,A}{\text{plím}\,B}$$

Ahora bien, a medida que el tamaño de la muestra aumenta indefinidamente, el segundo término tanto en el denominador como en el numerador de (5) tiende a cero (¿por qué?), lo cual genera

$$\text{plím}\,(\hat{\beta}_1) = \frac{\Pi_3}{\Pi_1} \tag{6}$$

lo que muestra que, aunque sea sesgado, $\hat{\beta}_1$ es un estimador consistente de $\beta_1$.

## 20.A.2 Estimación de los errores estándar de los estimadores de MC2E

El propósito de este apéndice es mostrar que los errores estándar de las estimaciones, obtenidos de la regresión de la segunda etapa del procedimiento de MC2E, utilizando la fórmula aplicable en la estimación por MCO, no son estimaciones "apropiadas" de los "verdaderos" errores estándar. Para ver esto, se utiliza el modelo del ingreso y oferta monetaria dado en las ecuaciones (20.4.1) y (20.4.2). Se estiman los parámetros de la función de oferta monetaria sobreidentificada a partir de la regresión de la segunda etapa como

$$Y_{2t} = \beta_{20} + \beta_{21}\hat{Y}_{1t} + u_t^* \qquad \textbf{(20.4.6)}$$

en donde

$$u_t^* = u_{2t} + \beta_{21}\hat{u}_t \qquad \textbf{(7)}$$

Ahora bien, cuando se efectúa la regresión (20.4.6), el error estándar de $\hat{\beta}_{21}$, por ejemplo, se obtiene de la siguiente expresión:

$$\text{var}\,(\hat{\beta}_{21}) = \frac{\hat{\sigma}_{u^*}^2}{\sum \hat{y}_{1t}^2} \qquad \textbf{(8)}$$

en donde

$$\hat{\sigma}_{u^*}^2 = \frac{\sum(\hat{u}_t^*)^2}{n-2} = \frac{\sum(Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}\hat{Y}_{1t})^2}{n-2} \qquad \textbf{(9)}$$

Pero $\sigma_{u^*}^2$ no es lo mismo que $\hat{\sigma}_{u_2}^2$, puesto que el último es una estimación insesgada de la verdadera varianza de $u_2$. Esta diferencia puede verificarse fácilmente en (7). Para obtener la verdadera $\hat{\sigma}_{u_2}^2$ (como se definió anteriormente), se procede de la siguiente manera:

$$\hat{u}_{2t} = Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}Y_{1t}$$

en donde $\hat{\beta}_{20}$ y $\hat{\beta}_{21}$ son las estimaciones de la regresión de la segunda etapa. Por tanto,

$$\hat{\sigma}_{u_2}^2 = \frac{\sum(Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}Y_{1t})^2}{n-2} \qquad \textbf{(10)}$$

Observe la diferencia entre (9) y (10): en (10) se utiliza la $Y_1$ observada más que la $Y_1$ estimada de la regresión de la primera etapa.

Después de estimar (10), la forma más fácil de corregir los errores estándar de los coeficientes estimados en la regresión de la segunda etapa es multiplicar cada uno de ellos por $\hat{\sigma}_{u_2}/\hat{\sigma}_{u\ddagger}$. Observe que si $Y_{1t}$ y $\hat{Y}_{1t}$ están muy cerca una de la otra, es decir, la $R^2$ en la regresión de la primera etapa es muy alta, el factor de corrección $\hat{\sigma}_{u_2}/\hat{\sigma}_{u^*}$ estará cercano a 1, en cuyo caso los errores estándar estimados en la regresión de la segunda etapa pueden considerarse como estimaciones verdaderas. En otras situaciones deberá utilizarse el factor de corrección anterior.

Capítulo 21

# Econometría de series de tiempo: algunos conceptos básicos

En el capítulo 1 mencionamos que uno de los dos tipos importantes de información para el análisis empírico lo conforman las **series de tiempo**. En este capítulo y el siguiente estudiaremos en detalle tal información, pues plantea diversos desafíos a econometristas y profesionales.

*En primer lugar*, el trabajo empírico basado en series de tiempo supone que la serie de tiempo en cuestión es **estacionaria**. Aunque en el capítulo 1 presentamos una idea intuitiva de estacionariedad, en este capítulo analizaremos este concepto con más profundidad. Más específicamente, trataremos de averiguar el significado de estacionariedad y la razón por la cual se debe tener en cuenta.

*En segundo lugar*, en el capítulo 12, sobre la autocorrelación, analizamos varias de sus causas. A veces la autocorrelación se origina porque las series de tiempo subyacentes no son estacionarias.

*En tercer lugar*, al efectuar la regresión de una variable de serie de tiempo sobre otra variable de serie de tiempo con frecuencia se obtiene una $R^2$ muy elevada (superior a 0.9) aunque no haya una relación significativa entre las dos. En ocasiones no se espera ninguna relación entre las dos variables; sin embargo, una regresión de una variable sobre la otra a menudo muestra una relación significativa. Esta situación ejemplifica el problema de la **regresión espuria,** o **disparatada**, cuya naturaleza analizaremos en breve. Por consiguiente, es muy importante averiguar si la relación entre las variables económicas es verdadera o espuria. En este capítulo veremos cómo aparece una regresión espuria cuando las series de tiempo no son estacionarias.

*En cuarto lugar,* algunas series de tiempo financieras, como los precios de las acciones, muestran lo que se conoce como **fenómeno de caminata aleatoria**. Lo anterior significa que la mejor predicción para el precio de una acción, por ejemplo de IBM, es igual a su precio actual más un choque puramente aleatorio (o término de error). De ser así, el pronóstico del precio de las acciones sería un ejercicio inútil.

*En quinto lugar*, los modelos de regresión que consideran series de tiempo son muy comunes para los pronósticos. En vista de lo expuesto, deseamos saber si tal pronóstico es válido cuando las series de tiempo sobre las cuales se basa no son estacionarias.

*Por último*, las pruebas de causalidad de Granger y Sims, analizadas en el capítulo 17, suponen que las series de tiempo del análisis son estacionarias. Por consiguiente, deben afectuarse antes las pruebas para la estacionariedad que las de causalidad.

Desde el principio, resulta necesaria una advertencia. El tema del análisis de las series de tiempo es muy amplio y siempre está en evolución; además, algunas matemáticas propias de las diversas técnicas del análisis de las series de tiempo son tan complejas que lo mejor que se puede

esperar de una obra introductoria como ésta es que proporcione un panorama general de algunos conceptos fundamentales. Para quienes deseen más información, se proporcionan referencias bibliográficas.[1]

## 21.1 Repaso rápido a una selección de series de tiempo económicas de Estados Unidos

Para empezar y dar al lector una idea de algunos conceptos un tanto esotéricos del análisis de series de tiempo de este capítulo, es útil considerar varias series de tiempo económicas de Estados Unidos de interés general. Estas series de tiempo son:

IPD = ingreso personal disponible real (miles de millones de dólares)
PIB = producto interno bruto (miles de millones de dólares)
GCP = gasto de consumo personal real (miles de millones de dólares)
UE = utilidades empresariales (miles de millones de dólares)
Dividendo = dividendos (miles de millones de dólares)

El periodo que abarcan estas cifras trimestrales es de I-1947 a IV-2007, para un total de 244 trimestres, y todos los datos se ajustaron por estacionalidad a la tasa anual. Todos los datos se recopilaron de FRED, el sitio Web del Banco de la Reserva Federal de San Luis sobre economía. El PIB, IPD y GCP se expresan en dólares constantes, en este caso dólares de 2000. Las UE y los dividendos se expresan en dólares nominales.

Para ahorrar espacio, los datos básicos se encuentran en el sitio Web del libro de texto. Sin embargo, para dar una idea de estos datos, trazamos las gráficas correspondientes en las siguientes dos figuras. La figura 21.1 es una gráfica de los datos de los logaritmos del PIB, IPD y GCP y en la figura 21.2 presentamos los logaritmos de las otras dos series de tiempo (UE y Dividendo). Es práctica común graficar el logaritmo de una serie de tiempo para tener una idea de la tasa de crecimiento de dicha serie. Una gráfica de los datos es por lo general el primer paso en el análisis de series de tiempo. En estas figuras, la letra L denota el logaritmo natural.

La primera impresión de estas dos figuras es que todas estas series de tiempo parecen "tender" hacia arriba, aunque con algunas fluctuaciones. Suponga que deseamos especular sobre la forma de estas curvas más allá del periodo muestral, por ejemplo, para todos los trimestres de 2008.[2] Esto es factible si se conoce el mecanismo estadístico, o estocástico, o el **proceso de generación de datos (PGD)** que dio origen a estas curvas. Pero ¿cuál es ese mecanismo? Para responder ésta y otras preguntas relacionadas es necesario estudiar cierto vocabulario "nuevo" que han definido los analistas de series de tiempo y que explicaremos a continuación.

[1] En un nivel introductorio, estas referencias pueden ser útiles: Gary Koop, *Analysis of Economic Data*, John Wiley & Sons, Nueva York, 2000; Jeff B. Cromwell, Walter C. Labys y Michel Terraza, *Univariate Test for Time Series Models*, Sage, California, Ansbury Park, 1994; Jeff B. Cromwell, Michael H. Hannan, Walter C. Labys y Michel Terraza, *Multivariate Tests for Time Series Models*, Sage, California, Ansbury Park, 1994; H.R. Seddighi, K.A. Lawler y A.V. Katos, *Econometrics: A Practical Approach*, Routledge, Nueva York, 2000. Para un nivel intermedio, véase Walter Enders, *Applied Econometric Time Series*, John Wiley & Sons, Nueva York, 1995; Kerry Patterson, *An Introduction to Applied Econometrics: A Time Series Approach*, St. Martin's Press, Nueva York, 2000; T.C. Mills, *The Econometric Modelling of Financial Time Series*, 2a. ed., Cambridge University Press, Nueva York, 1999; Marno Verbeek, *A Guide to Modern Econometrics*, John Wiley & Sons, Nueva York, 2000; Wojciech W. Charemza y Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling and Vector Autoregression*, 2a. ed., Edward Elgar Publisher, Nueva York, 1997. Para un nivel avanzado, consulte J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, Nueva Jersey, 1994; así como G.S. Maddala e In-Moo Kim, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998. En el nivel de aplicaciones, consulte B. Bhaskara Rao (ed.), *Cointegration for the Applied Economist*, St. Martin's Press, Nueva York, 1994, y Chandan Mukherjee, Howard White y Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, Nueva York, 1998.

[2] Desde luego, ya se tienen los datos reales para este periodo y se podrían comparar con los datos que se "predijeron" con base en el periodo anterior.

**FIGURA 21.1**
Logaritmos del PIB, IPD y GCP reales, Estados Unidos, 1947-2007 (por trimestre, miles de millones de dólares).

*Nota:* En la figura, la letra L denota el logaritmo natural.

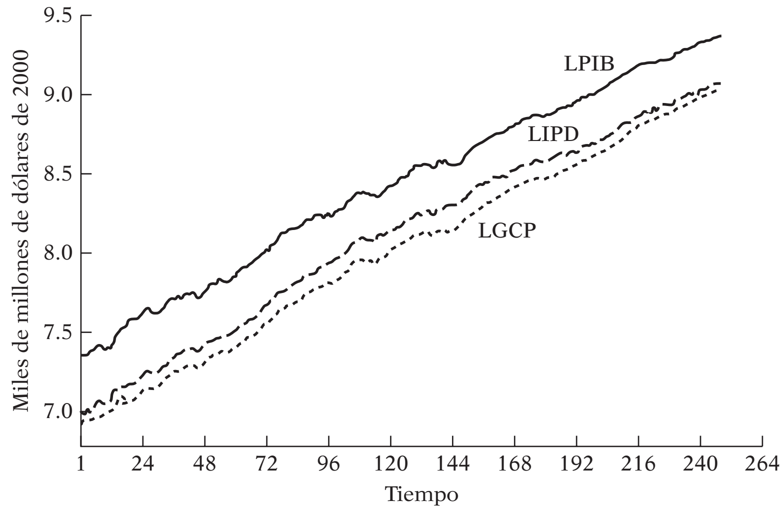
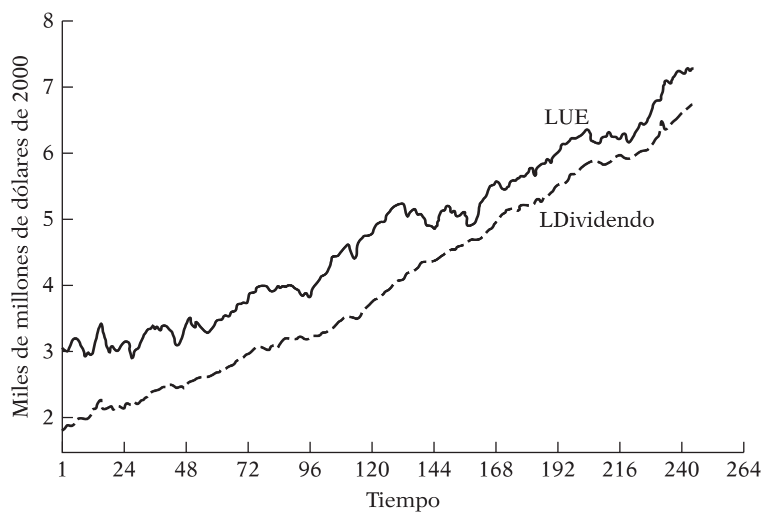



**FIGURA 21.2**
Logaritmos de utilidades empresariales (UE) y dividendos, Estados Unidos, 1947-2007 (por trimestre, miles de millones de dólares).

*Nota:* L denota logaritmo.




## 21.2 Conceptos fundamentales[3]

¿Cuál es este vocabulario? Consta de conceptos como:

1. Procesos estocásticos
2. Procesos estacionarios
3. Procesos puramente aleatorios
4. Procesos no estacionarios
5. Variables integradas
6. Modelos de caminata aleatoria
7. Cointegración
8. Tendencias deterministas y estocásticas
9. Pruebas de raíz unitaria

En las siguientes secciones analizaremos cada concepto. El análisis a menudo será heurístico. Siempre que sea posible, y útil, proporcionaremos ejemplos adecuados.

[3] El siguiente análisis se basa en Maddala *et al., op. cit.,* Charemza *et al., op. cit.,* y Carol Alexander, *Market Models: A Guide to Financial Data Analysis,* John Wiley & Sons, Nueva York, 2001.

## 21.3 Procesos estocásticos

*Un proceso estocástico o aleatorio es una colección de variables aleatorias ordenadas en el tiempo.*[4] Si $Y$ denota una variable aleatoria y es continua, se denota como $Y(t)$, pero si es discreta se expresa como $Y_t$. Un ejemplo del primer tipo es un electrocardiograma, y del segundo tipo, el PIB, IPD, etc. En vista de que la mayoría de los datos económicos se recopilan en puntos discretos de tiempo, para los propósitos de esta sección utilizaremos la notación $Y_t$ en vez de $Y(t)$. Si $Y$ representa al PIB, para los datos anteriores se tiene $Y_1, Y_2, Y_3, \ldots, Y_{242}, Y_{243}, Y_{244}$, donde el subíndice 1 denota la primera observación (es decir, el PIB del primer trimestre de 1947) y el subíndice 244 señala la última observación (es decir, el PIB del cuarto trimestre de 2007). *Tenga en cuenta que cada una de estas Y es una variable aleatoria.*

¿En qué sentido podemos considerar al PIB un proceso estocástico? Considere por ejemplo el PIB real de 3 759 997 millones de dólares del primer trimestre de 1970. En teoría, la cifra del PIB del primer trimestre de 1970 puede ser cualquier dígito, según el clima económico y político. La cifra 3 759 997 es una **realización** particular de todas esas posibilidades.[5] Por tanto, podemos decir que el PIB es un proceso estocástico y que los valores reales observados en el periodo del primer trimestre de 1947 al cuarto de 2007 son realizaciones particulares de ese proceso (es decir, una muestra). La distinción entre el proceso estocástico y su realización es semejante a la diferencia entre población y muestra en datos de corte transversal. De la misma forma como hacemos inferencias sobre la población a partir de datos muestrales, efectuamos inferencias sobre el proceso estocástico subyacente en las series de tiempo mediante la realización.

### Procesos estocásticos estacionarios

Un tipo de proceso estocástico que ha recibido gran atención y ha sido objeto de escrutinio por parte de los analistas de series de tiempo es el **proceso estocástico estacionario**. En términos generales, *se dice que un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza.* En la bibliografía sobre series de tiempo, un proceso estocástico como éste se conoce como **proceso estocástico débilmente estacionario**, **estacionario covariante**, **estacionario de segundo orden** o **proceso estocástico en amplio sentido**. Para efectos de este capítulo, y en la mayoría de las situaciones prácticas, basta este tipo de estacionariedad.[6]

Para explicar la estacionariedad débil, sea $Y_t$ una serie de tiempo estocástica con estas propiedades:

$$\textit{Media:} \qquad E(Y_t) = \mu \tag{21.3.1}$$

$$\textit{Varianza:} \qquad \operatorname{var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \tag{21.3.2}$$

$$\textit{Covarianza:} \qquad \gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] \tag{21.3.3}$$

donde $\gamma_k$, la covarianza (o autocovarianza) en el rezago $k$, es la covarianza entre los valores de $Y_t$ y $Y_{t+k}$, es decir, entre dos valores $Y$ separados $k$ periodos. Si $k = 0$, obtenemos $\gamma_0$, que es simple-

[4] El término "estocástico" proviene de la palabra griega "stokhos", que significa "blanco" u "objetivo". Si alguna vez ha jugado a los dardos con el propósito de atinarle al blanco, ¿cuántas veces acertó? De un centenar de tiros, quizá, si tuvo mucha suerte, le atinó al blanco unas cuantas veces; en las otras ocasiones, los dardos se esparcieron aleatoriamente alrededor del blanco.

[5] El valor de 3 759 997 millones de dólares se puede considerar el valor medio de todos los valores posibles del PIB para el primer trimestre de 1970.

[6] Una serie de tiempo es estrictamente estacionaria si *todos* los momentos de su distribución de probabilidad, y no sólo los dos primeros (es decir, la media y la varianza), son invariantes respecto del tiempo. Sin embargo, si el proceso estacionario es normal, el proceso estocástico débilmente estacionario también es estrictamente estacionario, pues el proceso estocástico normal está del todo especificado por sus dos momentos, la media y la varianza.

mente la varianza de $Y(=\sigma^2)$; si $k=1$, $\gamma_1$ es la covarianza entre dos valores adyacentes de $Y$, el tipo de covarianza encontrada en el capítulo 12 (recuerde el esquema autorregresivo de primer orden de Markov).

Suponga que el origen de $Y$ se desplaza de $Y_t$ a $Y_{t+m}$ (por ejemplo, del primer trimestre de 1947 al primero de 1952 de los datos del PIB). Ahora, si esperamos que $Y_t$ sea estacionaria, la media, la varianza y la covarianza de $Y_{t+m}$ deben ser las mismas que las de $Y_t$. *En resumen, si una serie de tiempo es estacionaria, su media, su varianza y su autocovarianza (en los diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto del tiempo.* Tal serie de tiempo tenderá a regresar a su media (llamada **reversión media**) y las fluctuaciones alrededor de esta media (medida por su varianza) tendrán una amplitud constante en términos generales.[7] Para decirlo de otro modo, un proceso estacionario no se desvía demasiado de su valor medio debido a la varianza finita. Como veremos enseguida, esto no ocurre con los procesos estocásticos no estacionarios. Debemos señalar que, en un proceso estacionario, la velocidad de la reversión media depende de las autocovarianzas: es rápida si las autocovarianzas son pequeñas y lenta cuando son grandes, como veremos en breve.

Si una serie de tiempo no es estacionaria en el sentido antes definido, se denomina **serie de tiempo no estacionaria** (recuerde que hablamos sólo de estacionariedad débil). En otras palabras, una serie de tiempo no estacionaria tendrá una *media que varía con el tiempo o una varianza que cambia con el tiempo, o ambas.*

¿Por qué las series de tiempo estacionarias son tan importantes? Porque si una serie de tiempo es no estacionaria, sólo podemos estudiar su comportamiento durante el periodo en consideración. Por tanto, cada conjunto de datos perteneciente a la serie de tiempo corresponderá a un episodio particular. En consecuencia, no es posible generalizar para otros periodos. Así, para propósitos de pronóstico, tales series de tiempo (no estacionarias) tienen poco valor práctico.

¿Cómo sabemos que una determinada serie de tiempo es estacionaria? En particular, ¿las series de tiempo de las figuras 21.1 y 21.2 son estacionarias? Analizaremos este tema importante en las secciones 21.8 y 21.9, cuando estudiemos varias pruebas para la estacionariedad. Pero, si juzgamos sólo con el sentido común, parece que las series de tiempo de las figuras 21.1 y 21.2 son no estacionarias, al menos en sus valores medios. Hablaremos de todo esto más adelante.

Antes de continuar, debemos mencionar un tipo especial de proceso estocástico (o de series de tiempo): el **proceso puramente aleatorio** o de **ruido blanco**. Se dice que un proceso es puramente aleatorio si tiene una media igual a cero, una varianza constante $\sigma^2$ y no está serialmente correlacionado.[8] Recordará que supusimos que el término de error $u_t$ que entra en el modelo clásico de regresión lineal —estudiado en la **parte 1** de este libro— era un proceso de ruido blanco denotado por $u_t \sim \text{IIDN}(0, \sigma^2)$; es decir, $u_t$ está independiente e idénticamente distribuido como una distribución normal con media cero y varianza constante. Este proceso se conoce como **proceso gaussiano de ruido blanco.**

## Procesos estocásticos no estacionarios

Aunque nuestro interés se centra en las series de tiempo estacionarias, a menudo se topa uno con series de tiempo no estacionarias, cuyo ejemplo clásico es el **modelo de caminata aleatoria** (MCA).[9] A menudo decimos que los precios de valores, como las acciones o las tasas de cambio, siguen una caminata aleatoria; es decir, son no estacionarios. Hay dos tipos de caminatas aleatorias: 1) caminata aleatoria sin deriva o sin desvío (es decir, sin término constante o de intercepto), y 2) caminata aleatoria con deriva o con desvío (es decir, hay un término constante).

[7] Esta observación pertenece a Keith Cuthbertson, Stephen G. Hall y Mark P. Taylor, *Applied Econometric Techniques*, The University of Michigan Press, p. 130.

[8] Si también es independiente, tal proceso se conoce como **estrictamente de ruido blanco**.

[9] El término "caminata aleatoria" a menudo se compara con el caminar de un borracho. Al dejar la cantina, el borracho se mueve una distancia aleatoria $u_t$ en el tiempo $t$ y continúa caminando de manera indefinida, con lo cual a la larga se aleja cada vez más de la cantina. Lo mismo se dice de los precios de las acciones. El precio de hoy de las acciones es igual al precio de ayer más un choque aleatorio.

### *Caminata aleatoria sin deriva*

Suponga que $u_t$ es un término de error de ruido blanco, con media 0 y varianza $\sigma^2$. Entonces decimos que la serie $Y_t$ es una caminata aleatoria si

$$Y_t = Y_{t-1} + u_t \tag{21.3.4}$$

En el modelo de caminata aleatoria, como se ve en (21.3.4), el valor de $Y$ en el tiempo $t$ es igual a su valor en el tiempo $(t-1)$ más un choque aleatorio; por tanto, es un modelo AR(1), en el lenguaje de los capítulos 12 y 17. Podemos pensar que (21.3.4) es una regresión de $Y$ en el tiempo $t$ sobre su valor rezagado un periodo. Los defensores de la **hipótesis del mercado de capital eficiente** argumentan que los precios de las acciones son en esencia aleatorios y, por tanto, no hay lugar para la especulación redituable en el mercado de valores: si se pudiese predecir el precio de las acciones del día siguiente con base en su precio del día anterior, todos seríamos millonarios.

Ahora bien, de (21.3.4), podemos escribir

$$Y_1 = Y_0 + u_1$$
$$Y_2 = Y_1 + u_2 = Y_0 + u_1 + u_2$$
$$Y_3 = Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3$$

En general, si el proceso comenzó en el tiempo 0 con un valor de $Y_0$, tenemos

$$Y_t = Y_0 + \sum u_t \tag{21.3.5}$$

Por tanto,

$$E(Y_t) = E\left(Y_0 + \sum u_t\right) = Y_0 \quad \text{(¿por qué?)} \tag{21.3.6}$$

De igual forma se demuestra que

$$\text{var}\,(Y_t) = t\sigma^2 \tag{21.3.7}$$

Como revelan las expresiones anteriores, la media de $Y$ es igual a su valor inicial (constante), pero conforme se incrementa $t$, su varianza aumenta de manera indefinida, lo que viola una condición de la estacionariedad. En resumen, el MCA sin deriva es un proceso estocástico no estacionario. En la práctica, $Y_0$ a menudo se iguala a cero, en cuyo caso $E(Y_t) = 0$.

Una característica importante del MCA es la *persistencia de los choques aleatorios* (es decir, los errores aleatorios), lo cual resulta evidente de (21.3.5): $Y_t$ es la suma de $Y_0$ inicial más la suma de los choques aleatorios. Como resultado, no se desvanece el impacto de un choque particular. Por ejemplo, si $u_2 = 2$, en vez de $u_2 = 0$, todas las $Y_t$ de $Y_2$ en adelante serán 2 unidades mayores, por lo que nunca cesa el efecto de este choque. Por esta razón decimos que la caminata aleatoria tiene *memoria infinita*. Como observa Kerry Patterson, la caminata aleatoria recuerda los choques por siempre;[10] es decir, tiene memoria infinita. La suma $\sum u_t$ se conoce también como **tendencia estocástica,** sobre la cual hablaremos en detalle más adelante.

Resulta interesante que si expresamos (21.3.4) como

$$(Y_t - Y_{t-1}) = \Delta Y_t = u_t \tag{21.3.8}$$

donde $\Delta$ es el operador de primeras diferencias, mismo que analizamos en el capítulo 12, resulta fácil probar que mientras que $Y_t$ es no estacionaria, sí lo es la serie de sus primeras diferencias. En otras palabras, las primeras diferencias de series de tiempo de caminata aleatoria son estacionarias. No obstante, hay más que decir al respecto.

[10] Kerry Patterson, *op. cit.*, capítulo 6.

*Caminata aleatoria con deriva*

Modifiquemos (21.3.4) de la siguiente forma:

$$Y_t = \delta + Y_{t-1} + u_t \tag{21.3.9}$$

donde $\delta$ se conoce como el **parámetro de deriva**. El término **deriva** proviene del hecho de que, si escribimos la ecuación anterior como

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \tag{21.3.10}$$

se demuestra que $Y_t$ se deriva o desvía hacia arriba o hacia abajo, según $\delta$ sea positiva o negativa. Observe que el modelo (21.3.9) también es un modelo AR(1).

Según el procedimiento analizado en la caminata aleatoria sin deriva, podemos demostrar que, para el modelo de caminata aleatoria con deriva (21.3.9),

$$E(Y_t) = Y_0 + t \cdot \delta \tag{21.3.11}$$

$$\text{var}(Y_t) = t\sigma^2$$

Como puede observar, para el MCA con deriva, la media, al igual que la varianza, se incrementa con el tiempo, lo que viola de nuevo las condiciones de la estacionariedad (débil). En resumen, el MCA, con o sin deriva, es un proceso estocástico no estacionario.

A fin de dar una ligera idea de la caminata aleatoria con y sin deriva, llevaremos a cabo dos simulaciones a continuación:

$$Y_t = Y_0 + u_t \tag{21.3.13}$$

donde $u_t$ son términos de error de ruido blanco de forma que cada $u_t \sim$ N(0, 1); es decir, cada $u_t$ sigue la distribución normal estándar. Mediante un generador de números aleatorios se obtuvieron 500 valores de $u$ y se generó $Y_t$ como se muestra en (21.3.13). Supusimos que $Y_0 = 0$. Por tanto, (21.3.13) es un MCA sin deriva.

Ahora considere

$$Y_t = \delta + Y_0 + u_t \tag{21.3.14}$$

que es un MCA sin deriva. Supusimos que los valores $u_t$ y $Y_0$ son como en (21.3.13) y que $\delta = 2$.

Las gráficas de los modelos (21.3.13) y (21.3.14) aparecen en las figuras 21.3 y 21.4, respectivamente. El lector puede comparar tales diagramas a la luz del análisis del MCA con y sin deriva.

**FIGURA 21.3**
Caminata aleatoria sin deriva.

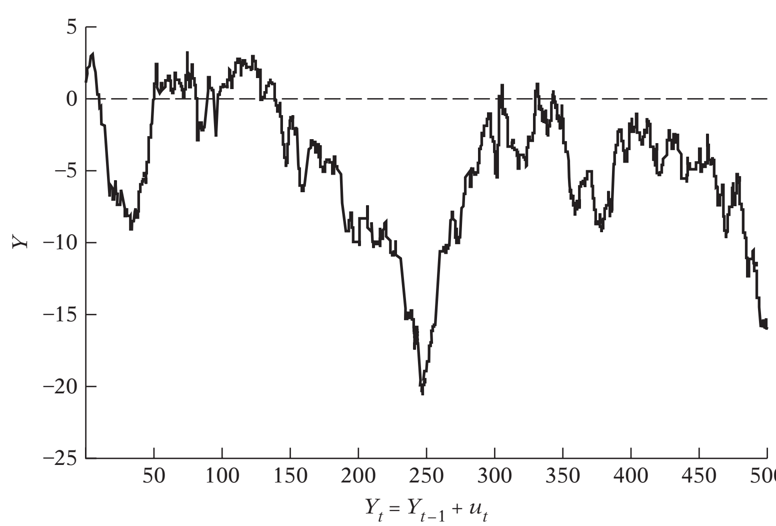
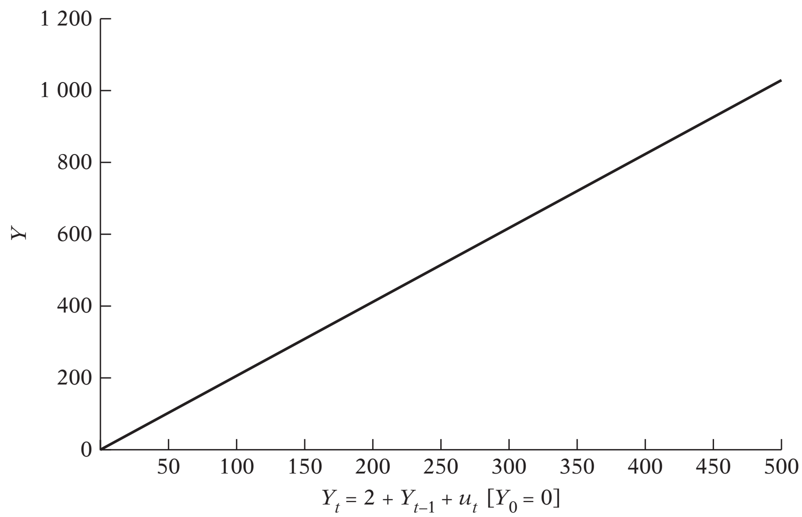


**FIGURA 21.4**
Caminata aleatoria con deriva.




El modelo de caminata aleatoria es un ejemplo de lo que se conoce en la bibliografía como **proceso de raíz unitaria**. Como este término es ya muy común en las referencias de series de tiempo, a continuación explicaremos lo que es un proceso de raíz unitaria.

## 21.4 Proceso estocástico de raíz unitaria

Escribimos el MCA (21.3.4) como:

$$Y_t = \rho Y_{t-1} + u_t \qquad -1 \leq \rho \leq 1 \tag{21.4.1}$$

Este modelo se parece al modelo autorregresivo de primer orden de Markov que analizamos en el capítulo de autocorrelación. Si $\rho = 1$, (21.4.1) se convierte en un MCA (sin deriva). Si $\rho$ es en efecto 1, tenemos lo que se conoce como **problema de raíz unitaria**; es decir, enfrentamos una situación de no estacionariedad. Ya sabemos que en este caso la varianza de $Y_t$ es no estacionaria. El nombre de raíz unitaria se debe a que $\rho = 1$.[11] *Por tanto, los términos* no estacionariedad, caminata aleatoria, raíz unitaria *y* tendencia estocástica *se consideran sinónimos.*

Sin embargo, si $|\rho| < 1$, es decir, si el valor absoluto de $\rho$ es menor que 1, podemos demostrar que la serie de tiempo $Y_t$ es estacionaria de acuerdo con la definición dada.[12]

Así, en la práctica, es importante averiguar si una serie de tiempo tiene una raíz unitaria.[13] En la sección 21.9 analizaremos varias pruebas de raíz unitaria, es decir, diversas pruebas para la estacionariedad. En dicha sección también determinaremos si las series de tiempo graficadas en las figuras 21.1 y 21.2 son estacionarias. Quizás el lector sospeche que no lo son. A su debido tiempo veremos esto.

[11] Una observación técnica: si $\rho = 1$, (21.4.1) se expresa como $Y_t - Y_{t-1} = u_t$. Ahora, con el **operador de rezago** $L$, de modo que $LY_t = Y_{t-1}$, $L^2Y_t = Y_{t-2}$, etc., (21.4.1) se escribe como $(1 - L)Y_t = u_t$. El término "raíz unitaria" se refiere a la raíz del polinomio en el operador de rezago. Si se tiene $(1 - L) = 0$, $L = 1$, de ahí el nombre de raíz unitaria.

[12] Si en (21.4.1) se supone que el valor inicial de $Y(= Y_0)$ es cero, $|\rho| < 1$ y $u_t$ es de ruido blanco, y tiene una distribución normal con una media cero y una varianza unitaria, por tanto se deduce que $E(Y_t) = 0$ y $(Y_t) = 1/(1 - \rho^2)$. Como ambas son constantes, por definición de estacionariedad débil, $Y_t$ es estacionaria. Por otra parte, como ya vimos, si $\rho = 1$, $Y_t$ es una caminata aleatoria o no estacionaria.

[13] Una serie de tiempo puede contener más de una raíz unitaria. Estudiaremos tal situación más adelante en este capítulo.

## 21.5 Procesos estocásticos estacionarios en tendencia (ET) y estacionarios en diferencias (ED)

La distinción entre procesos estocásticos (o series de tiempo) estacionarios y no estacionarios tiene una importancia fundamental para saber si la tendencia (la lenta evolución de largo plazo de la serie de tiempo en consideración) observada en las series de tiempo presentadas en las figuras 21.3 y 21.4 o en las series de tiempo económicas reales de las figuras 21.1 y 21.2 es **determinista** o **estocástica**. En términos generales, si la tendencia de una serie de tiempo es del todo predecible y no variable, se le llama **tendencia determinista**; si no es predecible, se le llama **tendencia estocástica**. Para formalizar la definición, considere el siguiente modelo de la serie de tiempo $Y_t$:

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \qquad \textbf{(21.5.1)}$$

donde $u_t$ es un término de error de ruido blanco y donde $t$ es el tiempo medido cronológicamente. Ahora tenemos las siguientes probabilidades:

**Caminata aleatoria pura:** Si en (21.5.1) $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 1$, obtenemos

$$Y_t = Y_{t-1} + u_t \qquad \textbf{(21.5.2)}$$

que no es otra cosa sino el MCA sin deriva y por tanto es no estacionario. Pero observe que si expresamos (21.5.2) como

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \qquad \textbf{(21.3.8)}$$

se convierte en estacionaria, como ya mencionamos. Por tanto, un MCA sin deriva es un **proceso estacionario en diferencias (PED)**.

**Caminata aleatoria con deriva:** Si en (21.5.1) $\beta_1 \neq 0$, $\beta_2 = 0$, $\beta_3 = 1$, obtenemos

$$Y_t = \beta_1 + Y_{t-1} + u_t \qquad \textbf{(21.5.3)}$$

que es una caminata aleatoria con deriva y en consecuencia es no estacionaria. Si la expresamos como

$$(Y_t - Y_{t-1}) = \Delta Y_t = \beta_1 + u_t \qquad \textbf{(21.5.3}a\textbf{)}$$

esto significa que $Y_t$ mostrará una tendencia positiva ($\beta_1 > 0$) o negativa ($\beta_1 < 0$) (figura 21.4). Tal tendencia se llama **tendencia estocástica**. La ecuación (21.5.3*a*) es un PED porque la no estacionariedad en $Y_t$ se elimina al tomar las primeras diferencias de las series de tiempo.

**Tendencia determinista:** Si en (21.5.1), $\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 = 0$, obtenemos

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad \textbf{(21.5.4)}$$

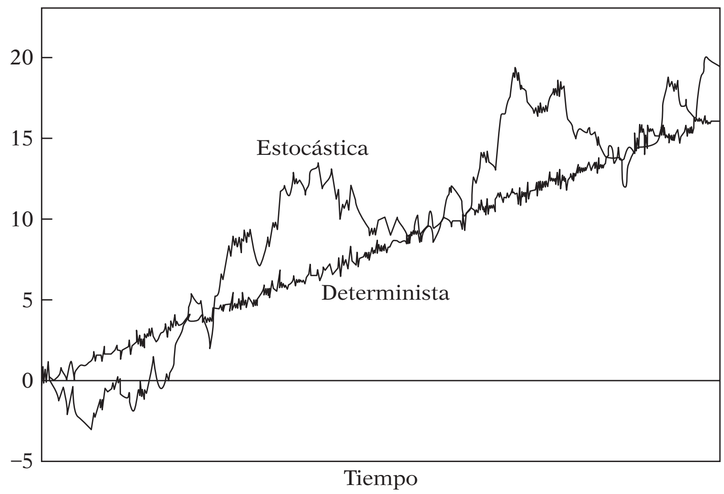
lo cual se llama **proceso estacionario en tendencia (PET)**. Aunque la media de $Y_t$ es $\beta_1 + \beta_2 t$ —no constante—, su varianza ($= \sigma^2$) sí lo es. Una vez que conocemos los valores de $\beta_1$ y $\beta_2$, podemos pronosticar la media sin ningún problema. Por tanto, si restamos la media de $Y_t$ de $Y_t$, la serie resultante será estacionaria; de ahí el nombre de **estacionario en tendencia**. Este procedimiento de eliminar la tendencia (determinista) se llama **supresión de tendencia**.

**Caminata aleatoria con deriva y tendencia determinista:** Si en (21.5.1) $\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 = 1$, obtenemos

$$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + u_t \qquad \textbf{(21.5.5)}$$

**FIGURA 21.5**
Tendencia determinista frente a tendencia estocástica.

Fuente: Charemza *et al., op. cit.*, p. 91.




en cuyo caso tenemos una caminata aleatoria con deriva y tendencia determinista, lo cual se aprecia si expresamos esta ecuación como

$$\Delta Y_t = \beta_1 + \beta_2 t + u_t \qquad \textbf{(21.5.5}\textbf{\textit{a}}\textbf{)}$$

que significa que $Y_t$ es no estacionaria.

**Tendencia determinista con componente estacionario AR(1):** Si en (21.5.1) $\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 < 1$, tenemos

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \qquad \textbf{(21.5.6)}$$

que es estacionaria alrededor de la tendencia determinista.

Para apreciar la diferencia entre una tendencia determinista y una estocástica, considere la figura 21.5.[14] La serie llamada estocástica en esta figura está generada por el MCA con deriva: $Y_t = 0.5 + Y_{t-1} + u_t$, donde se generaron 500 valores de $u_t$ a partir de la distribución estándar y donde el valor inicial de $Y$ se estableció como 1. La serie llamada determinista se genera de la siguiente forma: $Y_t = 0.5t + u_t$, donde $u_t$ se generó como antes y $t$ es el tiempo medido cronológicamente.

Como se ve a partir de la figura 21.5, en el caso de la tendencia determinista, las desviaciones de la línea de tendencia (que representa la media no estacionaria) son puramente aleatorias y se eliminan rápido; no contribuyen al desarrollo de largo plazo de las series de tiempo, el cual está determinado por el componente de la tendencia $0.5t$. En el caso de la tendencia estocástica, por otra parte, el componente aleatorio $u_t$ afecta el curso de largo plazo de la serie $Y_t$.

## 21.6 Procesos estocásticos integrados

El modelo de caminata aleatoria no es más que un caso específico de una clase más general de procesos estocásticos conocidos como **procesos integrados**. Recuerde que el MCA sin deriva es no estacionario, pero su serie de primeras diferencias, como se muestra en (21.3.8), es estacionaria. Por tanto, el MCA sin deriva se llama proceso **integrado de orden 1** y se denota como $I(1)$. De manera similar, si una serie de tiempo tiene que diferenciarse dos veces (es decir, se toman primeras diferencias de la serie de primeras diferencias) para hacerla estacionaria, esa serie de tiempo se denomina **integrada de orden 2**.[15] En general, si una serie de tiempo (no estacionaria)

[14] El siguiente análisis se basa en Wojciech W. Charemza *et al., op. cit.*, pp. 89-91.

[15] Por ejemplo, si $Y_t$ es $I(2)$, entonces $\Delta\Delta Y_t = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$ se convertirá en estacionaria. Pero observe que $\Delta\Delta Y_t = \Delta^2 Y_t \neq Y_t - Y_{t-2}$.

debe diferenciarse $d$ veces para hacerla estacionaria, decimos que la serie es **integrada de orden $d$**. Una serie de tiempo $Y_t$ integrada de orden $d$ se denota como $Y_t \sim I(d)$. Si una serie de tiempo es estacionaria desde el principio (es decir, si no requiere ninguna diferenciación), decimos que es integrada de orden cero y se denota mediante $Y_t \sim I(0)$. Por tanto, con los términos "serie de tiempo estacionaria" y "serie de tiempo integrada de orden cero" daremos a entender la misma cosa.

La mayoría de las series de tiempo económicas son $I(1)$; es decir, por lo general se convierten en estacionarias sólo después de tomar sus primeras diferencias. ¿Las series de tiempo mostradas en las figuras 21.1 y 21.2 son $I(1)$ o de orden mayor? Las examinaremos en las secciones 21.8 y 21.9.

### Propiedades de las series integradas

Podemos observar las siguientes propiedades de las series de tiempo integradas: sea $X_t$, $Y_t$ y $Z_t$ tres series de tiempo.

1. Si $X_t \sim I(0)$ y $Y_t \sim I(1)$, $Z_t = (X_t + Y_t) = I(1)$; es decir, una combinación lineal o suma de series de tiempo estacionaria y no estacionaria es no estacionaria.
2. Si $X_t \sim I(d)$, $Z_t = (a + bX_t) = I(d)$, donde $a$ y $b$ son constantes. Es decir, una combinación lineal de una serie $I(d)$ es también $I(d)$. Por tanto, si $X_t \sim I(0)$, $Z_t = (a + bX_t) \sim I(0)$.
3. Si $X_t \sim I(d_1)$ y $Y_t \sim I(d_2)$, $Z_t = (aX_t + bY_t) \sim I(d_2)$, donde $d_1 < d_2$.
4. Si $X_t \sim I(d)$ y $Y_t \sim I(d)$, $Z_t = (aX_t + bY_t) \sim I(d^*)$; $d^*$ es por lo general igual a $d$, pero en algunos casos $d^* < d$ (véase el tema de cointegración en la sección 21.11).

Como se ve por los enunciados anteriores, debemos poner especial atención al combinar dos o más series de tiempo que tengan diferente orden de integración.

Para ver la importancia de esto, considere el modelo de regresión de dos variables analizado en el capítulo 3, a saber, $Y_t = \beta_1 + \beta_2 X_t + u_t$. Según los supuestos clásicos de MCO, sabemos que

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{21.6.1}$$

donde las letras minúsculas, como siempre, indican la desviación de los valores medios. Suponga que $Y_t$ es $I(0)$ pero que $X_t$ es $I(1)$; es decir, la primera es estacionaria y la segunda no. Como $X_t$ es no estacionaria, su varianza se incrementará indefinidamente por tanto, domina el término del numerador en (21.6.1), con el resultado de que $\hat{\beta}_2$ convergirá a cero de manera asintótica (es decir, en muestras grandes) y no tendrá siquiera una distribución asintótica.[16]

## 21.7 El fenómeno de regresión espuria

Para ver por qué las series de tiempo estacionarias son tan importantes, considere los dos modelos de caminata aleatoria siguientes:

$$Y_t = Y_{t-1} + u_t \tag{21.7.1}$$

$$X_t = X_{t-1} + v_t \tag{21.7.2}$$

donde se generaron 500 observaciones de $u_t$ a partir de $u_t \sim N(0, 1)$ y 500 observaciones de $v_t$ a partir de $v_t \sim N(0, 1)$, además de que se supuso que los valores iniciales de $Y$ y $X$ eran cero. También se supuso que $u_t$ y $v_t$ no están serial ni mutuamente correlacionadas. Como ya sabemos, ambas series de tiempo son no estacionarias; es decir, son $I(1)$ o exhiben tendencias estocásticas.

[16] Esta observación se debe a Maddala *et al., op. cit.*, p. 26.

Suponga que hacemos la regresión de $Y_t$ sobre $X_t$. Como $Y_t$ y $X_t$ son procesos no correlacionados $I(1)$, $R^2$ de la regresión de $Y$ sobre $X$ debe tender a cero; es decir, no debe haber ninguna relación entre las dos variables. Pero vea los resultados de la regresión:

| Variable | Coeficiente | Error estándar | Estadístico *t* |
|---|---|---|---|
| *C* | −13.2556 | 0.6203 | −21.36856 |
| *X* | 0.3376 | 0.0443 | 7.61223 |
| | $R^2$ = 0.1044 | *d* = 0.0121 | |

Como puede observar, el coeficiente de $X$ es muy significativo estadísticamente, y aunque el valor de $R^2$ es bajo, es estadísticamente distinto de cero. A partir de estos resultados, uno estaría tentado a concluir que existe una relación estadística significativa entre $Y$ y $X$, aunque *a priori* se pensara que no habría ninguna. Lo anterior resume el **fenómeno de regresión espuria** o **regresión sin sentido** descubierto por Yule,[17] quien mostró además que la correlación (espuria) puede persistir en las series de tiempo no estacionarias aunque la muestra sea muy grande. Que hay algo malo en la regresión anterior lo indica el valor extremadamente bajo de la $d$ de Durbin-Watson, el cual indica una autocorrelación muy fuerte de primer orden. De acuerdo con Granger y Newbold, *$R^2 > d$ es una buena regla práctica para sospechar que la regresión estimada es espuria,* como en el ejemplo anterior. Podemos añadir que la $R^2$ y el estadístico $t$ de dicha regresión espuria son engañosos y que los estadísticos $t$ no están distribuidos como la distribución $t$ (de Student) y, por tanto, no se pueden probar con ellos hipótesis sobre los parámetros.

Que los resultados de la regresión presentados antes carezcan de sentido se advierte con facilidad al hacer la regresión de las primeras diferencias de $Y_t$ $(= \Delta Y_t)$ sobre las primeras diferencias de $X_t$ $(= \Delta X_t)$; recuerde que aunque $Y_t$ y $X_t$ son no estacionarias, sus primeras diferencias sí lo son. En esta regresión veremos que $R^2$ es prácticamente cero, como debe ser, y que la $d$ de Durbin-Watson es de casi 2. En el ejercicio 21.24 se le pedirá realizar esta regresión y verificar el enunciado anterior.

Aunque resulta drástico, este ejemplo es un recordatorio de que debemos tener mucho cuidado al llevar a cabo un análisis de regresión basado en series de tiempo que exhiban tendencias estocásticas. Así, hay que tomar muchas precauciones al interpretar de más los resultados de la regresión basados en variables $I(1)$. Por ejemplo, vea el ejercicio 21.26. En alguna medida, lo anterior resulta verdadero para las series de tiempo sujetas a tendencias deterministas, de lo cual se da un ejemplo en el ejercicio 21.25.

## 21.8 Pruebas de estacionariedad

A estas alturas es probable que el lector tenga una buena idea sobre la naturaleza de los procesos estocásticos estacionarios y su importancia. En la práctica se enfrentan dos preguntas importantes: 1) ¿Cómo sabemos si una serie de tiempo determinada es estacionaria? 2) Si tenemos que una serie de tiempo determinada es no estacionaria, ¿hay alguna forma de que se convierta en estacionaria? Abordaremos la primera pregunta en este apartado y la segunda en la sección 21.10.

Antes de proceder, tenga en cuenta que sobre todo nos interesa la estacionariedad débil o covarianza.

Aunque hay varias pruebas para la estacionariedad, sólo analizamos las que se estudian de manera prominente en la bibliografía. En esta sección examinaremos dos pruebas: 1) el análisis

[17] G.U. Yule, "Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series", en *Journal of the Royal Statistical Society*, vol. 89, 1926, pp. 1-64. Para amplias simulaciones Monte Carlo sobre regresiones espurias, véase C.W.J. Granger y P. Newbold, "Spurious Regressions in Econometrics", en *Journal of Econometrics*, vol. 2, 1974, pp. 111-120.

gráfico y 2) la prueba del correlograma. Debido a la importancia que le otorgamos en el pasado reciente, en el siguiente apartado estudiaremos la *prueba de raíz unitaria*. Ilustramos las pruebas mencionadas con ejemplos adecuados.

## 1. Análisis gráfico

Como ya mencionamos, antes de efectuar una prueba formal, siempre es aconsejable graficar la serie de tiempo en estudio, como se hizo en las figuras 21.1 y 21.2 con los datos de series de tiempo sobre indicadores económicos de Estados Unidos que se presentan en el sitio Web del libro de texto. Estas gráficas proporcionan una pista inicial respecto de la posible naturaleza de las series de tiempo. Por ejemplo, considere la serie de tiempo PIB de la figura 21.1. Observará que, a lo largo del periodo de estudio, el logaritmo del PIB se incrementó; es decir, muestra una tendencia ascendente, lo cual deja entrever que quizá esté variando la media del logaritmo del PIB. Esto tal vez indique que la serie logarítmica del PIB es no estacionaria, lo cual es más o menos verdadero para las otras series de tiempo económicas de Estados Unidos de la figura 21.2. Esa intuición es el comienzo de una prueba más formal de estacionariedad.

## 2. Función de autocorrelación (FAC) y correlograma

Una prueba sencilla de estacionariedad se basa en la denominada **función de autocorrelación (FAC)**. La FAC en el rezago $k$, denotada por $\rho_k$, se define como

$$\begin{aligned}\rho_k &= \frac{\gamma_k}{\gamma_0} \\ &= \frac{\text{covarianza en el rezago } k}{\text{varianza}}\end{aligned} \tag{21.8.1}$$

donde la covarianza en el rezago $k$ y la varianza son como se definieron anteriormente. Observe que si $k = 0$, $\rho_0 = 1$ (¿por qué?).

Como la covarianza y la varianza se miden en las mismas unidades, $\rho_k$ es *un número sin unidad de medida, o puro*. Se encuentra entre $-1$ y $+1$, igual que cualquier coeficiente de correlación. Si graficamos $\rho_k$ respecto de $k$, la gráfica obtenida se conoce como **correlograma poblacional**.

Como, en la práctica, sólo tenemos una realización de un proceso estocástico (es decir, la muestra), sólo podemos calcular la **función de autocorrelación muestral**, $\hat{\rho}_k$. Para tal efecto, debemos calcular primero la **covarianza muestral** en el rezago $k$, $\hat{\gamma}_k$, y la **varianza muestral**, $\hat{\gamma}_0$ definidas como[18]

$$\hat{\gamma}_k = \frac{\sum(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n} \tag{21.8.2}$$

$$\hat{\gamma}_0 = \frac{\sum(Y_t - \bar{Y})^2}{n} \tag{21.8.3}$$

donde $n$ es el tamaño de la muestra y $\bar{Y}$ es la media muestral.

Por consiguiente, la función de autocorrelación muestral en el rezago $k$ es

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \tag{21.8.4}$$

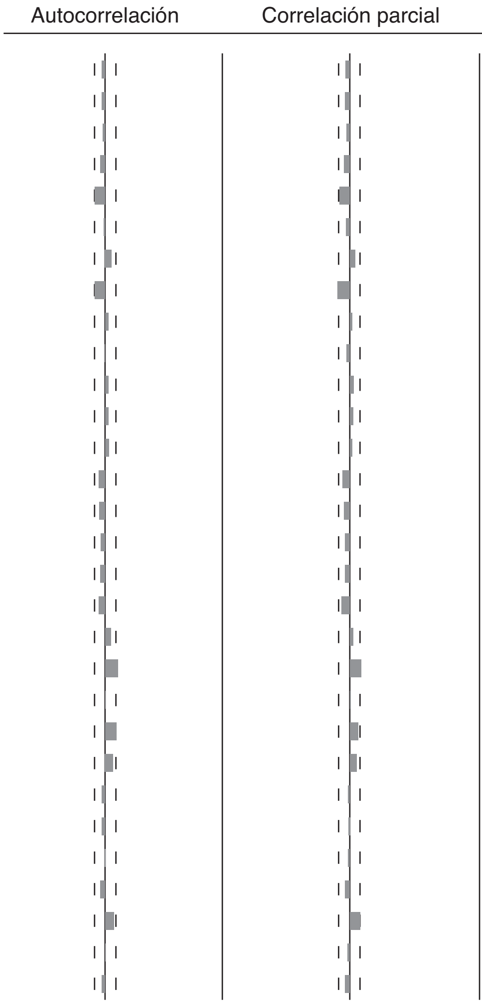
que es simplemente la razón entre la covarianza muestral (en el rezago $k$) y la varianza muestral. La gráfica de $\hat{\rho}_k$ frente a $k$ se conoce como **correlograma muestral**.

¿Cómo saber con un correlograma si una serie de tiempo particular es estacionaria? Para este propósito, primero presentaremos correlogramas muestrales de un proceso puramente aleatorio

[18] En sentido estricto, debemos dividir la covarianza muestral en el rezago $k$ por $(n - k)$ y la varianza muestral entre $(n - 1)$ en vez de hacerlo entre $n$ (¿por qué?), en donde $n$ es el tamaño de la muestra.

**FIGURA 21.6**
Correlograma del término de error de ruido blanco *u*. AC = autocorrelación, ACP = autocorrelación parcial (capítulo 22), Est. Q = estadístico *Q*, Prob = Probabilidad.

Muestra: 2 500
Observaciones incluidas: 499



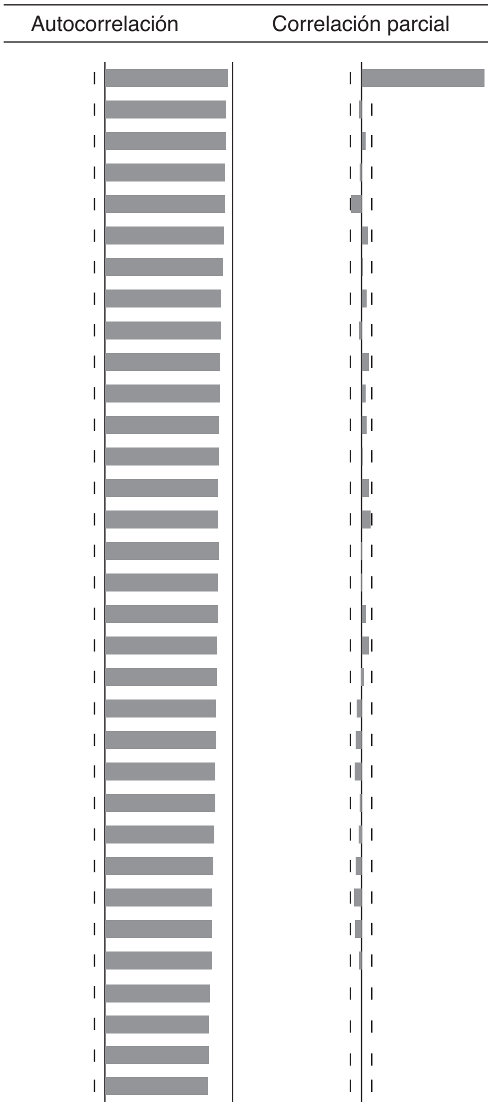

| | AC | ACP | Est. Q | Prob |
|---|---|---|---|---|
| 1 | –0.022 | –0.022 | 0.2335 | 0.629 |
| 2 | –0.019 | –0.020 | 0.4247 | 0.809 |
| 3 | –0.009 | –0.010 | 0.4640 | 0.927 |
| 4 | –0.031 | –0.031 | 0.9372 | 0.919 |
| 5 | –0.070 | –0.072 | 3.4186 | 0.636 |
| 6 | –0.008 | –0.013 | 3.4493 | 0.751 |
| 7 | 0.048 | 0.045 | 4.6411 | 0.704 |
| 8 | –0.069 | –0.070 | 7.0385 | 0.532 |
| 9 | 0.022 | 0.017 | 7.2956 | 0.606 |
| 10 | –0.004 | –0.011 | 7.3059 | 0.696 |
| 11 | 0.024 | 0.025 | 7.6102 | 0.748 |
| 12 | 0.024 | 0.027 | 7.8993 | 0.793 |
| 13 | 0.026 | 0.021 | 8.2502 | 0.827 |
| 14 | –0.047 | –0.046 | 9.3726 | 0.806 |
| 15 | –0.037 | –0.030 | 10.074 | 0.815 |
| 16 | –0.026 | –0.031 | 10.429 | 0.843 |
| 17 | –0.029 | –0.024 | 10.865 | 0.863 |
| 18 | –0.043 | –0.050 | 11.807 | 0.857 |
| 19 | 0.038 | 0.028 | 12.575 | 0.860 |
| 20 | 0.099 | 0.093 | 17.739 | 0.605 |
| 21 | 0.001 | 0.007 | 17.739 | 0.665 |
| 22 | 0.065 | 0.060 | 19.923 | 0.588 |
| 23 | 0.053 | 0.055 | 21.404 | 0.556 |
| 24 | –0.017 | –0.004 | 21.553 | 0.606 |
| 25 | –0.024 | –0.005 | 21.850 | 0.644 |
| 26 | –0.008 | –0.008 | 21.885 | 0.695 |
| 27 | –0.036 | –0.027 | 22.587 | 0.707 |
| 28 | 0.053 | 0.072 | 24.068 | 0.678 |
| 29 | –0.004 | –0.011 | 24.077 | 0.725 |
| 30 | –0.026 | –0.025 | 24.445 | 0.752 |

de ruido blanco y un proceso de caminata aleatoria. Regresemos al MCA sin deriva (21.3.13). Ahí generamos una muestra de 500 términos de error, las *u,* a partir de la distribución normal estandarizada. El correlograma para estos 500 términos de error puramente aleatorios es como se muestra en la figura 21.6; se muestran en este correlograma hasta 30 rezagos. En breve comentaremos cómo elegir la longitud del rezago.

Por el momento, sólo observe la columna AC, que es la función de autocorrelación muestral, y el primer diagrama de la izquierda, llamado autocorrelación. La línea vertical continua de este diagrama representa el eje cero; las observaciones por arriba de esta línea son valores positivos, y los que están por debajo, negativos. Como resulta evidente a partir de este diagrama, para un proceso puramente de ruido blanco, las autocorrelaciones en distintos rezagos se ubican alrededor del cero. *Ésta es una imagen de un correlograma de una serie de tiempo estacionaria.* Por tanto, si el correlograma de una serie de tiempo real (económica) se parece al correlograma de una serie de tiempo de ruido blanco, podemos decir que dicha serie de tiempo es quizá estacionaria.

**FIGURA 21.7**
Correlograma de una serie de tiempo de caminata aleatoria. Véanse las definiciones en la figura 21.6.

Muestra: 2 500
Observaciones incluidas: 499



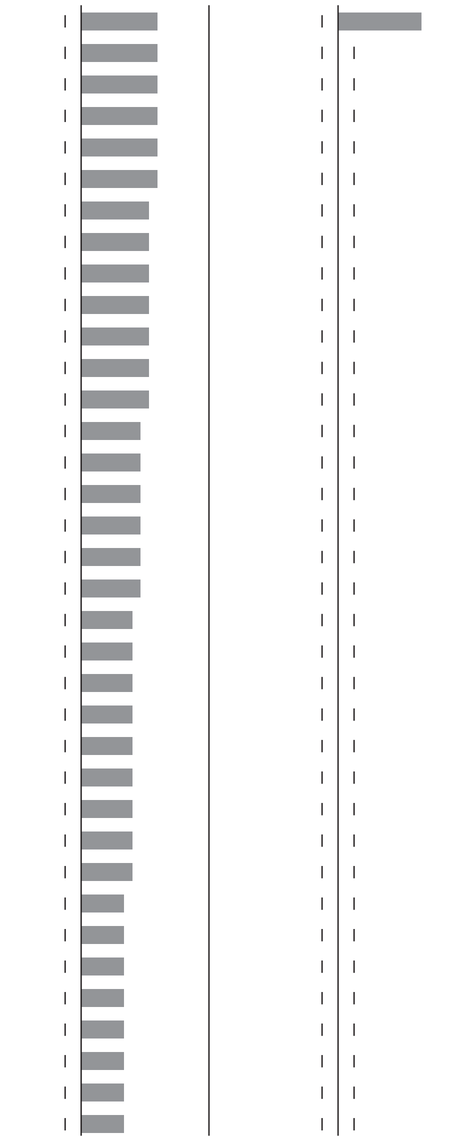
| | AC | ACP | Est. Q | Prob |
|---|---|---|---|---|
| 1 | 0.992 | 0.992 | 493.86 | 0.000 |
| 2 | 0.984 | 0.000 | 980.68 | 0.000 |
| 3 | 0.976 | 0.030 | 1 461.1 | 0.000 |
| 4 | 0.969 | 0.005 | 1 935.1 | 0.000 |
| 5 | 0.961 | –0.059 | 2 402.0 | 0.000 |
| 6 | 0.953 | 0.050 | 2 862.7 | 0.000 |
| 7 | 0.946 | 0.004 | 3 317.3 | 0.000 |
| 8 | 0.939 | 0.040 | 3 766.4 | 0.000 |
| 9 | 0.932 | –0.009 | 4 210.1 | 0.000 |
| 10 | 0.927 | 0.055 | 4 649.1 | 0.000 |
| 11 | 0.921 | 0.018 | 5 083.9 | 0.000 |
| 12 | 0.916 | 0.039 | 5 514.9 | 0.000 |
| 13 | 0.912 | 0.002 | 5 942.4 | 0.000 |
| 14 | 0.908 | 0.056 | 6 367.0 | 0.000 |
| 15 | 0.905 | 0.061 | 6 789.8 | 0.000 |
| 16 | 0.902 | 0.000 | 7 210.6 | 0.000 |
| 17 | 0.899 | 0.006 | 7 629.4 | 0.000 |
| 18 | 0.896 | 0.030 | 8 046.7 | 0.000 |
| 19 | 0.894 | 0.053 | 8 463.1 | 0.000 |
| 20 | 0.892 | 0.013 | 8 878.7 | 0.000 |
| 21 | 0.890 | –0.041 | 9 292.6 | 0.000 |
| 22 | 0.886 | –0.040 | 9 704.1 | 0.000 |
| 23 | 0.882 | –0.044 | 10 113. | 0.000 |
| 24 | 0.878 | –0.012 | 10 518. | 0.000 |
| 25 | 0.873 | –0.023 | 10 920. | 0.000 |
| 26 | 0.867 | –0.041 | 11 317 | 0.000 |
| 27 | 0.860 | –0.055 | 11 709. | 0.000 |
| 28 | 0.853 | –0.045 | 12 095. | 0.000 |
| 29 | 0.846 | –0.010 | 12 476. | 0.000 |
| 30 | 0.839 | 0.008 | 12 851. | 0.000 |
| 31 | 0.832 | –0.006 | 13 221. | 0.000 |
| 32 | 0.825 | 0.003 | 13 586. | 0.000 |
| 33 | 0.819 | –0.006 | 13 946. | 0.000 |

Ahora observe el correlograma de una serie de caminata aleatoria como se genera por (21.3.13). La gráfica se muestra en la figura 21.7. La característica más sobresaliente de este correlograma es que los coeficientes de autocorrelación, para diversos rezagos, son muy altos, incluso hasta para un rezago de 33 trimestres. De hecho, si consideramos rezagos de hasta 60 trimestres, los coeficientes de autocorrelación son muy altos; en el rezago 60, el coeficiente es de casi 0.7. La figura 21.7 es un correlograma habitual de una serie de tiempo no estacionaria. El coeficiente de autocorrelación comienza en un nivel muy alto y disminuye de modo muy lento hacia cero, conforme se prolonga el rezago.

Consideremos un ejemplo concreto. Examinemos el correlograma de la serie de tiempo LPIB graficada con base en los datos de series de tiempo económicas de Estados Unidos del sitio Web del libro (sección 21.1). En la figura 21.8 se muestra el correlograma de hasta 36 rezagos. El correlograma de hasta 36 rezagos del LPIB también muestra un patrón semejante al del corre-

**FIGURA 21.8**
Correlograma del LPIB de Estados Unidos, I-1947 a IV-2007. Véanse las definiciones en la figura 21.6.

Muestra: I-1947 IV-2007
Observaciones incluidas: 244




| | AC | ACP | Est. Q | Prob |
|---|---|---|---|---|
| 1 | 0.977 | 0.977 | 235.73 | 0.000 |
| 2 | 0.954 | –0.009 | 461.43 | 0.000 |
| 3 | 0.931 | –0.010 | 677.31 | 0.000 |
| 4 | 0.908 | –0.006 | 883.67 | 0.000 |
| 5 | 0.886 | –0.003 | 1 080.9 | 0.000 |
| 6 | 0.864 | –0.001 | 1 269.3 | 0.000 |
| 7 | 0.843 | –0.006 | 1 449.3 | 0.000 |
| 8 | 0.822 | –0.006 | 1 621.0 | 0.000 |
| 9 | 0.801 | –0.010 | 1 784.6 | 0.000 |
| 10 | 0.780 | –0.004 | 1 940.6 | 0.000 |
| 11 | 0.759 | –0.007 | 2 089.0 | 0.000 |
| 12 | 0.738 | –0.013 | 2 230.0 | 0.000 |
| 13 | 0.718 | 0.003 | 2 364.1 | 0.000 |
| 14 | 0.699 | –0.005 | 2 491.5 | 0.000 |
| 15 | 0.679 | –0.001 | 2 612.4 | 0.000 |
| 16 | 0.660 | –0.004 | 2 727.2 | 0.000 |
| 17 | 0.642 | –0.002 | 2 836.2 | 0.000 |
| 18 | 0.624 | 0.002 | 2 939.6 | 0.000 |
| 19 | 0.607 | 0.003 | 3 037.8 | 0.000 |
| 20 | 0.590 | –0.003 | 3 130.9 | 0.000 |
| 21 | 0.573 | –0.003 | 3 219.3 | 0.000 |
| 22 | 0.557 | –0.003 | 3 303.1 | 0.000 |
| 23 | 0.541 | –0.001 | 3 382.5 | 0.000 |
| 24 | 0.526 | 0.007 | 3 457.9 | 0.000 |
| 25 | 0.511 | 0.002 | 3 529.4 | 0.000 |
| 26 | 0.496 | –0.005 | 3 597.2 | 0.000 |
| 27 | 0.482 | –0.011 | 3 661.4 | 0.000 |
| 28 | 0.467 | –0.009 | 3 722.0 | 0.000 |
| 29 | 0.453 | –0.005 | 3 779.2 | 0.000 |
| 30 | 0.438 | –0.006 | 3 833.1 | 0.000 |
| 31 | 0.424 | –0.005 | 3 883.9 | 0.000 |
| 32 | 0.411 | 0.004 | 3 931.6 | 0.000 |
| 33 | 0.398 | 0.004 | 3 976.7 | 0.000 |
| 34 | 0.385 | –0.001 | 4 019.1 | 0.000 |
| 35 | 0.373 | –0.009 | 4 058.9 | 0.000 |
| 36 | 0.360 | –0.010 | 4 096.3 | 0.000 |

lograma del modelo de caminata aleatoria de la figura 21.7. El coeficiente de autocorrelación comienza con un valor muy alto en el rezago 1 (0.977) y disminuye muy lentamente. Por tanto, parece que la serie de tiempo PIB es no estacionaria. Si graficamos los correlogramas de otras series de tiempo económicas de Estados Unidos de la figura 21.1 y 21.2 observaremos patrones similares, lo cual lleva a la conclusión de que todas estas series de tiempo son no estacionarias; tal vez sean no estacionarias respecto de la media o la varianza, o ambas.

Aquí podemos abordar dos cuestiones prácticas. En primer lugar, ¿cómo elegir la longitud del rezago para calcular la FAC?, y en segundo, ¿cómo determinar si un coeficiente de autocorrelación es estadísticamente significativo en un cierto rezago? A continuación damos las respuestas.

*Elección de la longitud del rezago*

Se trata básicamente de un asunto empírico. Una regla práctica es calcular la FAC hasta un tercio o una cuarta parte de la longitud de la serie de tiempo. En vista de que para los datos económicos de este ejemplo tenemos 244 observaciones trimestrales, según esta regla, los rezagos de 61 a 81 trimestres servirán. Para ahorrar espacio, sólo mostramos 36 rezagos en la gráfica de la FAC en la figura 21.8. El mejor consejo práctico es comenzar con rezagos lo bastante grandes y luego reducirlos mediante un criterio estadístico, como *el criterio de información Akaike o de Schwarz*, que analizamos en el capítulo 13. Por otra parte, también podemos utilizar cualquiera de las siguientes pruebas.

## Significancia estadística de los coeficientes de autocorrelación

Considere, por ejemplo, el correlograma de la serie de tiempo LPIB de la figura 21.8. ¿Cómo decidir si el coeficiente de correlación, 0.780, en el rezago 10 (trimestres) es estadísticamente significativo? La significancia estadística de cualquier $\hat{\rho}_k$ se juzga mediante su error estándar. Bartlett demostró que si una serie de tiempo es puramente aleatoria, es decir, si es una muestra de ruido blanco (figura 21.6), los coeficientes de autocorrelación muestrales $\hat{\rho}_k$ son *aproximadamente*[19]

$$\hat{\rho}_k \sim N(0, 1/n) \tag{21.8.5}$$

es decir, en muestras grandes, los coeficientes de autocorrelación muestrales están normalmente distribuidos y tienen media cero y varianza igual a 1 sobre el tamaño de la muestra. Como hay 244 observaciones, la varianza es $1/244 \approx 0.0041$, y el error estándar, $\sqrt{0.0041} \approx 0.0640$. Por tanto, según las propiedades de la distribución normal estándar, el intervalo de confianza de 95% para cualquier (población) $\rho_k$ es:

$$\hat{\rho}_k \pm 1.96(0.0640) = \hat{\rho}_k \pm 0.1254 \tag{21.8.6}$$

En otras palabras,

$$\text{Prob}\,(\hat{\rho}_k - 0.1254 \leq \rho_k \leq \hat{\rho}_k + 0.1254) = 0.95 \tag{21.8.7}$$

Si el intervalo anterior incluye el valor cero, no rechazamos la hipótesis de que la verdadera $\rho_k$ es cero, pero si este intervalo no incluye 0, rechazamos la hipótesis de que la verdadera $\rho_k$ es cero. Al aplicar esto al valor estimado de $\hat{\rho}_{10} = 0.780$, el lector puede verificar que el intervalo de confianza de 95% para la verdadera $\rho_{10}$ es $(0.780 \pm 0.1254)$ o $(0.6546, 0.9054)$.[20] Es obvio que este intervalo no incluye el valor cero, lo cual indica que hay 95% de confianza de que la verdadera $\rho_{10}$ sea significativamente diferente de cero.[21] Como se ve, incluso en el rezago 20 la $\rho_{20}$ es estadísticamente significativa en un nivel de 5%.

En lugar de probar la significancia estadística de cualquier coeficiente de autocorrelación individual, para probar la *hipótesis conjunta* de que todos los $\rho_k$ hasta ciertos rezagos son simultáneamente iguales a cero, podemos utilizar el **estadístico *Q*** desarrollado por Box y Pierce, que se define como[22]

$$Q = n \sum_{k=1}^{m} \hat{\rho}_k^2 \tag{21.8.8}$$

---

[19] M.S. Bartlett, "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series", en *Journal of the Royal Statistical Society*, serie B, vol. 27, 1946, pp. 27-41.

[20] El tamaño de la muestra de 244 observaciones es razonablemente grande para usar la aproximación normal.

[21] Por otra parte, si divide el valor estimado de cualquier $\rho_k$ entre el error estándar $(\sqrt{1/n})$ para una $n$ lo bastante grande, obtendrá el valor estándar $Z$, cuya probabilidad se obtiene fácilmente a partir de la tabla normal estándar. Por tanto, para el valor estimado de $\rho_{10} = 0.780$, el valor $Z$ es $0.780/0.1066 = 7.32$ (aproximadamente). Si la verdadera $\rho_{10}$ fuera en efecto cero, la probabilidad de obtener un valor $Z$ igual o mayor que 7.32 es muy pequeña, por lo que rechazamos la hipótesis de que la verdadera $\rho_{10}$ es cero.

[22] G.E. P. Box y D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, vol. 65, 1970, pp. 1509-1526.

donde $n$ = tamaño de la muestra y $m$ = longitud del rezago. El estadístico $Q$ es común para probar si una serie de tiempo es de ruido blanco. En muestras grandes, este estadístico se distribuye *aproximadamente* como la distribución ji cuadrada con $m$ gl. En una aplicación, si la $Q$ calculada excede el valor $Q$ crítico de la distribución ji cuadrada en el nivel de significancia seleccionado, podemos rechazar la hipótesis nula de que todos los $\rho_k$ (verdaderos) son iguales a cero; por lo menos algunos de ellos deben ser diferentes de cero.

Una variante del estadístico $Q$ de Box-Pierce es el **estadístico Ljung-Box (LB)**, que se define como[23]

$$\text{LB} = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \sim \chi^2 m \tag{21.8.9}$$

Aunque en muestras grandes tanto el estadístico $Q$ como el estadístico LB siguen la distribución ji cuadrada con $m$ gl, se ha visto que el estadístico LB tiene mejores propiedades en muestras pequeñas (más potente, en el sentido estadístico) que el estadístico $Q$.[24]

De regreso al ejemplo del LPIB de la figura 21.8, el valor del estadístico $Q$ hasta el rezago 36 es cercano a 4 096. La probabilidad de obtener tal valor de $Q$ según la hipótesis nula de que la suma de los 36 cuadrados de los coeficientes de autocorrelación estimados sea cero es prácticamente nula, como lo muestran las cifras de la última columna. Por consiguiente, la conclusión es que la serie de tiempo LPIB probablemente es no estacionaria, con lo cual se refuerza la conjetura basada en la figura 21.1: la serie LPIB tal vez era no estacionaria. En el ejercicio 21.16 se pide al lector confirmar que las otras cuatro series de tiempo económicas de Estados Unidos también son no estacionarias.

## 21.9 Prueba de raíz unitaria

Otra prueba sobre estacionariedad (o no estacionariedad) que se populariza cada vez más se conoce como **prueba de raíz unitaria**. Primero la explicaremos, luego la ilustraremos y después consideraremos algunas limitantes de esta prueba.

El punto de partida es el proceso (estocástico) de raíz unitaria que vimos en la sección 21.4. Se inicia con

$$Y_t = \rho Y_{t-1} + u_t \qquad -1 \le \rho \le 1 \tag{21.4.1}$$

donde $u_t$ es un término de error de ruido blanco.

Sabemos que si $\rho = 1$, es decir, en el caso de la raíz unitaria, (21.4.1) se convierte en un modelo de caminata aleatoria sin deriva, del cual sabemos también que es un proceso estocástico no estacionario. Por consiguiente, ¿por qué no simplemente hacer la regresión de $Y_t$ sobre su valor rezagado (de un periodo) $Y_{t-1}$ y se averigua si la $\rho$ estimada es estadísticamente igual a 1? De ser así, $Y_t$ es no estacionaria. Ésta es la idea general de la prueba de raíz unitaria para la estacionariedad.

Sin embargo, no podemos estimar la ecuación (21.4.1) por MCO y probar la hipótesis de que $\rho = 1$ por medio de la prueba $t$ acostumbrada, porque esa prueba tiene un sesgo muy marcado en el caso de una raíz unitaria. Por tanto, manipulamos (21.4.1) de la siguiente forma: restamos $Y_{t-1}$ de ambos miembros de la ecuación (21.4.1) para obtener:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ &= (\rho - 1)Y_{t-1} + u_t \end{aligned} \tag{21.9.1}$$

la cual también se expresa como:

$$\Delta Y_t = \delta Y_{t-1} + u_t \tag{21.9.2}$$

donde $\delta = (\rho - 1)$ y $\Delta$, como siempre, es el operador de primeras diferencias.

[23] G.M. Ljung y G.P.E. Box, "On a Measure of Lack of Fit in Time Series Models", en *Biometrika*, vol. 66, 1978, pp. 66-72.

[24] Los estadísticos $Q$ y LB tal vez no resulten apropiados en todos los casos. Para conocer una crítica de lo anterior, consulte Maddala *et al., op. cit.*, p. 19.

Por tanto, en la práctica, en vez de estimar (21.4.1), calculamos (21.9.2) y probamos la hipótesis (nula) de que $\delta = 0$, y la hipótesis alternativa es que $\delta < 0$ (nota 25). Si $\delta = 0$, entonces $\rho = 1$; es decir, tenemos una raíz unitaria, lo cual significa que la serie de tiempo en consideración es no estacionaria.

Antes de proceder con la estimación de (21.9.2) debemos observar que si $\delta = 0$, entonces (21.9.2) se convertirá en

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \tag{21.9.3}$$

Como $u_t$ es un término de error de ruido blanco, entonces es estacionario, lo cual significa que las primeras diferencias de una serie de tiempo de caminata aleatoria son estacionarias, una observación que ya habíamos hecho.

Ahora reconsideremos la estimación de (21.9.2). Esto es muy simple: sólo hay que tomar las primeras diferencias de $Y_t$ y hacer la regresión sobre $Y_{t-1}$, a fin de ver si el coeficiente estimado de la pendiente en esta regresión ($= \hat{\delta}$) es o no cero. Si es cero, concluimos que $Y_t$ es no estacionaria; pero si es negativa, se infiere que $Y_t$ es estacionaria.[25] La única interrogante es saber con qué prueba averiguar si el coeficiente estimado de $Y_{t-1}$ en (21.9.2) es o no cero. Uno estaría tentado a utilizar la prueba *t* usual. Por desgracia, según la hipótesis nula de que $\delta = 0$ (es decir, $\rho = 1$), el valor *t* del coeficiente estimado de $Y_{t-1}$ no sigue la distribución *t* ni siquiera en muestras grandes, es decir, no tiene una distribución normal asintótica.

¿Cuál es la alternativa? Dickey y Fuller probaron que según la hipótesis nula de que $\delta = 0$, el valor estimado *t* del coeficiente $Y_{t-1}$ en (21.9.2) sigue el **estadístico $\tau$ (tau)**.[26] Estos autores calcularon los valores críticos del *estadístico tau* con base en simulaciones Monte Carlo. Una muestra de esos valores críticos se da en el **apéndice D**, tabla D.7. La tabla es limitada, pero MacKinnon preparó tablas más extensas, ya incorporadas en diferentes software estadísticos.[27] En la bibliografía, el **estadístico** o **prueba tau** se conoce como **prueba Dickey-Fuller (DF)**, en honor a sus descubridores. Resulta interesante que si rechazamos la hipótesis de que $\delta = 0$ (es decir, la serie de tiempo es estacionaria), podemos utilizar la prueba *t* (de Student) usual. Tenga en cuenta que la prueba Dickey-Fuller es unidireccional porque la hipótesis alternativa es que $\delta < 0$ (o $\rho < 1$).

El procedimiento real para aplicar la prueba DF supone diversas decisiones. Al analizar la naturaleza del proceso de raíz unitaria en las secciones 21.4 y 21.5 observamos que un proceso de caminata aleatoria tal vez no tiene deriva, o quizá sí, o posiblemente tiene tendencia determinista y estocástica. A fin de permitir las distintas posibilidades, la prueba DF se estima en tres diferentes formas, es decir, conforme a tres hipótesis nulas:

**$Y_t$ es una caminata aleatoria:** $\Delta Y_t = \delta Y_{t-1} + u_t$ **(21.9.2)**

**$Y_t$ es una caminata aleatoria con deriva:** $\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$ **(21.9.4)**

**$Y_t$ es una caminata aleatoria con deriva alrededor de una tendencia determinista:** $\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$ **(21.9.5)**

---

[25] Porque $\delta = (\rho - 1)$, por lo que la estacionariedad $\rho$ debe ser menor que uno. Para que esto suceda, $\delta$ debe ser negativa.

[26] D.A. Dickey y W.A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", en *Journal of the American Statistical Association*, vol. 74, 1979, pp. 427-431. Véase también W.A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, Nueva York, 1976.

[27] J.G. MacKinnon, "Critical Values of Cointegration Test", en R.E. Engle y C.W.J. Granger (eds.), *Long-Run Economic Relationships: Readings in Cointegration*, cap. 13, Oxford University Press, Nueva York, 1991.

donde $t$ es la variable de tiempo o de tendencia. En cada caso, las hipótesis son:

*Hipótesis nula:* $H_0: \delta = 0$ (es decir, existe una raíz unitaria, la serie de tiempo es no estacionaria o tiene tendencia estocástica).

*Hipótesis alternativa*: $H_1: \delta < 0$ (es decir, la serie de tiempo es estacionaria, posiblemente alrededor de una tendencia determinista).[28]

Si rechazamos la hipótesis nula, esto significa que 1) $Y_t$ es estacionaria con media cero en el caso de la ecuación (21.9.2) o que 2) $Y_t$ es estacionaria con una media distinta de cero en el caso de (21.9.4). En el caso de la ecuación (21.9.5), podemos probar que $\delta < 0$ (es decir, no hay tendencia estocástica) y $\alpha \neq 0$ (es decir, la existencia de una tendencia determinista) simultáneamente, mediante la prueba $F$ pero con los valores críticos tabulados por Dickey y Fuller. Cabe señalar que una serie de tiempo puede contener tanto una tendencia estocástica como una determinista.

*Es extremadamente importante observar que los valores críticos de la prueba tau para probar la hipótesis de que $\delta = 0$ son diferentes en cada una de las tres especificaciones anteriores de la prueba DF*, lo cual se ve claramente en el **apéndice D**, tabla D.7. Es más, si, por ejemplo, la especificación (21.9.4) es correcta pero se estima (21.9.2), cometemos un error de especificación, cuyas consecuencias ya conocemos desde el capítulo 13. La misma regla se aplica si estimamos (21.9.4) en vez del verdadero (21.9.5). Desde luego, no hay forma de saber cuál especificación es la correcta. Resulta inevitable hacer pruebas de ensayo y error, no obstante la minería de datos.

El procedimiento real de estimación es el siguiente: Estimamos (21.9.2), (21.9.3) o (21.9.4) mediante MCO; dividimos el coeficiente estimado de $Y_{t-1}$ en cada caso entre su error estándar a fin de calcular el estadístico tau ($\tau$) y consultamos las tablas DF (o cualquier software estadístico). Si el valor absoluto calculado del estadístico tau ($|\tau|$) excede la DF *absoluta* o los valores críticos tau de MacKinnon, rechazamos la hipótesis de que $\delta = 0$, en cuyo caso la serie de tiempo es estacionaria. Por otra parte, si el $|\tau|$ calculado no excede el valor crítico tau, no rechazamos la hipótesis nula, en cuyo caso la serie de tiempo es no estacionaria. Hay que asegurarse de utilizar los valores críticos $\tau$ apropiados. En la mayoría de las aplicaciones, el valor tau es negativo. Por consiguiente, también vale decir que si el valor tau calculado (negativo) es más pequeño (es decir, más negativo) que el valor crítico tau, rechazamos la hipótesis nula (es decir, la serie de tiempo es estacionaria); de lo contrario, no la rechazamos (es decir, la serie de tiempo es no estacionaria).

Regresemos al ejemplo de las series de tiempo del PIB de Estados Unidos. Para estas series, los resultados de las tres regresiones (21.9.2), (21.9.4) y (21.9.5) son los siguientes: la variable dependiente en cada caso es $\Delta Y_t = \Delta \text{LPIB}_t$, donde LPIB es el logaritmo del PIB real.

$$\begin{aligned} \widehat{\Delta \text{LPIB}}_t &= 0.000968\text{LPIB}_{t-1} \\ t &= (12.9270) \qquad R^2 = 0.0147 \qquad d = 1.3194 \end{aligned} \tag{21.9.6}$$

$$\begin{aligned} \widehat{\Delta \text{LPIB}}_t &= 0.0221 - 0.00165\text{LPIB}_{t-1} \\ t &= (2.4342) \quad (-1.5294) \qquad R^2 = 0.0096 \qquad d = 1.3484 \end{aligned} \tag{21.9.7}$$

$$\begin{aligned} \widehat{\Delta \text{LPIB}}_t &= 0.2092 + 0.0002t - 0.0269\text{LPIB}_{t-1} \\ t &= (1.8991) \quad (1.7040) \quad (-1.8102) \\ & \qquad R^2 = 0.0215 \qquad d = 1.3308 \end{aligned} \tag{21.9.8}$$

[28] Descartamos la posibilidad de que $\delta > 0$ porque en ese caso $\rho > 1$, y de ser así, la serie de tiempo subyacente sería explosiva.

El principal interés en todas estas regresiones radica en el valor $t(= \tau)$ del coeficiente $\text{LPIB}_{t-1}$. Si analizamos la tabla D.7 del **apéndice D**, observaremos que los valores críticos tau a 5% para un tamaño de muestra de 250 (el número más próximo a la muestra de 244 observaciones que estudiamos aquí) son $-1.95$ (sin intercepto, sin tendencia), $-2.88$ (intercepto pero sin tendencia) y $-3.43$ (intercepto y tendencia). *EViews* y otros paquetes estadísticos proporcionan valores críticos para el tamaño de muestra del análisis.

Antes de examinar los resultados, tenemos que decidir cuál de los tres modelos es el adecuado. Debemos descartar el modelo (21.9.6) porque el coeficiente $\text{LPIB}_{t-1}$, que es igual a $\delta$, es positivo. Pero en vista de que $\delta = (\rho - 1)$, una $\delta$ positiva implicaría que $\rho > 1$. Aunque es una posibilidad teórica, se descarta en este caso porque la serie de tiempo LPIB sería explosiva.[29] Por tanto, no quedan más que los modelos (21.9.7) y (21.9.8). En ambos casos, el coeficiente estimado $\delta$ es negativo, lo cual implica que la $\rho$ estimada es menor que 1. Para ambos modelos, los valores estimados $\rho$ son 0.9984 y 0.9731, respectivamente. Ahora, la única pregunta pendiente es saber si estos valores son estadísticamente menores que 1 de manera significativa, para que podamos decir que la serie de tiempo del PIB es estacionaria.

Para el modelo (21.9.7), el valor $\tau$ estimado es $-1.5294$, mientras que el valor crítico $\tau$ a 5%, como ya señalamos, es $-2.88$. Como en términos *absolutos* el primer valor es más pequeño que el segundo, la conclusión es que la serie de tiempo LPIB es no estacionaria.[30]

Sucede lo mismo con el modelo (21.9.8). El valor $\tau$ calculado de $-1.8102$, en términos absolutos, es menor incluso que el valor crítico a 5% de $-3.43$.

Por tanto, con base en el análisis gráfico, el correlograma y la prueba Dickey-Fuller, la conclusión es que para los periodos trimestrales de 1947 a 2007, la serie de tiempo LPIB de Estados Unidos fue no estacionaria; es decir, contenía una raíz unitaria, o tenía una tendencia estocástica.

## La prueba Dickey-Fuller aumentada (DFA)

Al llevar a cabo la prueba DF en (21.9.2), (21.9.4) o (21.9.5) supusimos que el término de error $u_t$ no estaba correlacionado. Pero Dickey y Fuller desarrollaron una prueba cuando dicho término sí está correlacionado, la cual se conoce como **prueba Dickey-Fuller aumentada (DFA)**. Esta prueba implica "aumentar" las tres ecuaciones anteriores mediante la adición de los valores rezagados de la variable dependiente $\Delta Y_t$. Para un ejemplo específico, suponga que utilizamos (21.9.5). La prueba DFA consiste en este caso en estimar la siguiente regresión:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \qquad \textbf{(21.9.9)}$$

donde $\varepsilon_t$ es un término de error puro de ruido blanco y donde $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, $\Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$, etc. El número de términos de diferencia rezagados que debemos incluir con frecuencia se determina de manera empírica, con la idea de incluir los términos suficientes para que el término de error en (21.9.9) no esté serialmente relacionado y sea posible obtener una estimación insesgada de $\delta$, el coeficiente de $Y_{t-1}$ rezagado. *EViews 6* tiene una opción que selecciona automáticamente la longitud del rezago con base en los criterios de información de Akaike, Schwarz y otros. En la DFA se sigue probando $\delta = 0$, y además esta prueba sigue la misma distribución asintótica que el estadístico DF, por lo que se sirven los mismos valores críticos.

Con el fin de dar una idea general de este procedimiento estimamos (21.9.9) para la serie LPIB. Como se tienen datos trimestrales, decidimos usar cuatro rezagos. Los resultados de la regresión DFA fueron los siguientes:[31]

---

[29] De manera más técnica, como (21.9.2) es una ecuación diferencial de primer orden, la llamada condición de estabilidad requiere que $|\rho| < 1$.

[30] Otra forma de expresar esto sería que el valor $\tau$ calculado deba ser más negativo que el valor crítico $\tau$, lo cual no sucede aquí. Por tanto, mantenemos la conclusión. Como en general se espera que $\delta$ sea negativa, el estadístico estimado $\tau$ tendrá signo negativo. Por tanto, un valor $\tau$ grande y negativo suele ser un indicio de estacionariedad.

[31] Se consideraron diferencias rezagadas de orden superior, pero fueron insignificantes.

$$\widehat{\Delta \text{LPIB}_t} = 0.2677 + 0.0003t - 0.0352\text{LPIB}_{t-1} + 0.2990\Delta\text{LPIB}_{t-1} + 0.1451\Delta\text{LPIB}_{t-2} - 0.0621\Delta\text{LPIB}_{t-3} - 0.0876\Delta\text{LPIB}_t$$
$$t = (2.4130) \quad (2.2561) \quad (-2.3443) \qquad (4.6255) \qquad (2.1575) \qquad (-0.9205) \qquad (-1.3438)$$
$$R^2 = 0.1617 \qquad d = 2.0075 \tag{21.9.10}$$

El valor $t(=\tau)$ del coeficiente $\text{LPIB}_{t-1}$ rezagado ($=\delta$) es $-2.3443$, que en términos absolutos es incluso mucho menor que el valor crítico $\tau$ a 10% de $-3.1378$, lo cual indica de nuevo que aun después de tener cuidado de la posible autocorrelación en el término de error, la serie LPIB es no estacionaria. (*Nota:* El comando @trend de *EViews* genera automáticamente la variable de tiempo o tendencia.)

¿Puede ser éste el resultado de haber elegido sólo cuatro valores rezagados de ΔLPIB? Apliquemos el criterio de Schwartz con 14 valores rezagados de ΔLPIB, lo que arrojó el valor tau $\delta$ de $-1.8102$. Aun entonces, este valor tau no fue significativo en el nivel de 10% (el valor crítico tau en este nivel fue de $-3.1376$). Al parecer, el logaritmo del PIB es no estacionario.

## Prueba de la significancia de más de un coeficiente: prueba *F*

Suponga que estimamos el modelo (21.9.5) y probamos la hipótesis de que $\beta_1 = \beta_2 = 0$, es decir, el modelo es MCA sin deriva ni tendencia. Para probar esta hipótesis conjunta utilizamos la prueba *F restringida* analizada en el capítulo 8. Es decir, estimamos (21.9.5) (la regresión no restringida) y luego estimamos (21.9.5) otra vez, lo que elimina el intercepto y la tendencia. Luego utilizamos la prueba *F* restringida, como se muestra en la ecuación (8.6.9), excepto que no se emplea la tabla *F* convencional a fin de obtener los valores críticos *F*. Como hicieron para el estadístico $\tau$, Dickey y Fuller desarrollaron valores críticos *F* para esta situación; una muestra de lo anterior se da en el **apéndice D**, tabla D.7. En el ejercicio 21.27 se proporciona un ejemplo de lo anterior.

## Las pruebas de raíz unitaria Phillips-Perron (PP)[32]

Un supuesto importante de la prueba DF es que los términos de error $u_t$ están idéntica e independientemente distribuidos. La prueba DFA ajusta la prueba DF a fin de tener cuidado de una posible correlación serial en los términos de error al agregar los términos de diferencia rezagados de la regresada. Phillips y Perron utilizan *métodos estadísticos no paramétricos* para evitar la correlación serial en los términos de error, sin añadir términos de diferencia rezagados. Como la distribución asintótica de la prueba PP es la misma que la prueba DFA, no examinaremos con mayor detalle este tema.

## Prueba de cambios estructurales

Los datos macroeconómicos introducidos en la sección 21.1 (consulte los datos reales en el sitio Web del libro) corresponden al periodo 1947-2007, 61 años. En este periodo la economía de Estados Unidos pasó por varios ciclos económicos de diferentes duraciones. Los ciclos económicos están marcados por periodos de recesiones y de expansiones. Es muy probable que un ciclo económico sea distinto de otro, lo que puede reflejar **rupturas estructurales** o **cambios estructurales** en la economía.

Por ejemplo, considere el primer embargo petrolero, en 1973. Los precios del petróleo se cuadriplicaron. Los precios volvieron a aumentar de manera sustancial después del segundo embargo petrolero, en 1979. Como es natural, estas conmociones afectan el comportamiento económico. Por tanto, si queremos hacer una regresión del gasto de consumo personal (GCP) sobre el ingreso personal disponible (IPD), es muy probable que el intercepto, la pendiente o ambas varíen de un ciclo económico a otro (recuerde la prueba de Chow de rupturas estructurales). Esto es lo que se entiende por cambios estructurales.

[32] P.C.B. Phillips y P. Perron, "Testing for a Unit Root in Time Series Regression", en *Biometrika*, vol. 75, 1988, pp. 335-346. La prueba PP ahora se incluye en varios software estadísticos.

Por ejemplo, Perron sostiene que las pruebas estándar de la hipótesis de raíz unitaria pueden no ser confiables en presencia de cambios estructurales.[33] Existen varias formas de probar los cambios estructurales y explicarlos; la más sencilla supone el uso de variables dicótomas. Sin embargo, un análisis a fondo de las diversas pruebas de rupturas estructurales va mucho más allá del texto y es mejor dejarlo a las referencias.[34] No obstante, vea el ejercicio 21.28.

## Crítica de las pruebas de raíz unitaria[35]

Se han analizado varias pruebas de raíz unitaria y además existen todavía otras más. La pregunta es: ¿por qué hay tantas pruebas de raíz unitaria? La respuesta radica en su **tamaño** y **potencia**. Por tamaño de la prueba nos referimos al nivel de significancia (es decir, la probabilidad de cometer un error tipo I), y por potencia de una prueba a la probabilidad de rechazar la hipótesis nula cuando es falsa. Calculamos la potencia de una prueba al restar la probabilidad de un error tipo II de 1; el error tipo II es la probabilidad de aceptar una hipótesis nula falsa. El máximo poder es 1. Casi todas las pruebas de raíz unitaria se basan en la hipótesis nula de que la serie de tiempo que se analiza tiene una raíz unitaria; o sea, es no estacionaria. La hipótesis alterna es que la serie de tiempo es estacionaria.

### *Tamaño de la prueba*

Recordará, del capítulo 13, la distinción entre los niveles de significancia nominales y los verdaderos. La prueba DF es sensible a la forma en que se lleva a cabo. Recuerde que analizamos tres variedades de pruebas DF: 1) una caminata puramente aleatoria, 2) una caminata aleatoria con deriva y 3) una caminata aleatoria con deriva y tendencia. Si, por ejemplo, el verdadero modelo es 1) pero se estima un modelo 2) y se concluye que, por ejemplo, con un nivel de significancia de 5% la serie es estacionaria, esta conclusión puede ser errónea porque el verdadero nivel de significancia en este caso es mucho mayor que 5%.[36] El tamaño de la distorsión también puede deberse a la exclusión de componentes de promedios móviles (PM) del modelo (sobre promedios móviles, véase el capítulo 22).

### *Potencia de la prueba*

La mayoría de las pruebas del tipo DF tienen poco poder; es decir, tienden a aceptar la nulidad de la raíz unitaria con más frecuencia de la garantizada. En otras palabras, estas pruebas pueden encontrar una raíz unitaria aunque no exista. Hay varias razones para esto. *En primer lugar*, la potencia depende del *lapso* de los datos más que del solo tamaño de la muestra. Para una muestra dada de tamaño $n$, la potencia es mayor cuando el lapso es grande. En consecuencia, la(s) prueba(s) basada(s) en 30 observaciones sobre un lapso de 30 años quizá tengan más potencia que una basada por ejemplo en 100 observaciones durante un lapso de 100 días. *En segundo lugar*, si $\rho \approx 1$ pero no es exactamente 1, la prueba de raíz unitaria puede diagnosticar la serie de tiempo como no estacionaria. *En tercer lugar*, estos tipos de prueba suponen una raíz unitaria; es decir, suponen que la serie de tiempo dada es $I(1)$. Pero si una serie de tiempo es integrada de orden mayor que 1, por ejemplo, $I(2)$, habrá más de una raíz unitaria. De ser así, se puede utilizar la **prueba Dickey-Pantula**.[37] *En cuarto lugar*, si hay rupturas estructurales en una serie de tiempo (véase el capítulo sobre variables dicótomas) debidas, por ejemplo, al embargo petrolero por parte de la OPEP, las pruebas de raíz unitarias quizá no las reflejen.

---

[33] P. Perron, "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica,* vol. 57, 1989, pp. 1361-1401.

[34] Hay un análisis accesible en James H. Stock y Mark W. Watson, *Introduction to Econometrics,* 2a. ed., Pearson/Addison-Wesley, Boston, 2007, pp. 565-571. Para un análisis más minucioso, véase G.S. Maddala e In-Moo Kim, *Unit Roots, Cointegration, and Structural Change,* Cambridge University Press, Nueva York, 1998.

[35] Para un análisis detallado, véase Terrence C. Mills, *op. cit.*, pp. 87-88.

[36] Para un experimento Monte Carlo al respecto, véase Charemza *et al., op. cit.*, p. 114.

[37] D.A. Dickey y S. Pantula, "Determining the Order of Differencing in Autoregressive Processes", en *Journal of Business and Economic Statistics*, vol. 5, 1987, pp. 455-461.

Por tanto, al aplicar las pruebas de raíz unitaria se deben tener en cuenta sus limitaciones. Desde luego, Perron y Ng, Elliot, Rothenberg y Stock, Fuller y Leybounre[38] modificaron esas pruebas. Debido a lo anterior, Maddala y Kim afirman que las pruebas tradicionales DF, DFA y PP deben descartarse. Quizá eso llegue a suceder conforme los paquetes de software de econometría incorporen nuevas pruebas. Pero debemos añadir que hasta la fecha no existe una prueba uniformemente poderosa de la hipótesis de la raíz unitaria.

## 21.10 Transformación de las series de tiempo no estacionarias

Ahora que conocemos el problema asociado a las series de tiempo no estacionarias, surge la pregunta práctica de qué hay que hacer. Para evitar el problema de la regresión espuria que pudiese surgir al hacer la regresión de una serie de tiempo no estacionaria sobre una o más series de tiempo no estacionarias tenemos que transformar las series de tiempo no estacionarias en estacionarias. El método de transformación depende de que las series de tiempo sean procesos estacionarios en diferencias (PED) o procesos estacionarios con tendencia (PET). Consideraremos cada caso a su debido tiempo.

### Procesos estacionarios en diferencias

Si una serie de tiempo tiene una raíz unitaria, las primeras diferencias de tales series son estacionarias.[39] En consecuencia, la solución aquí es tomar las primeras diferencias de las series de tiempo.

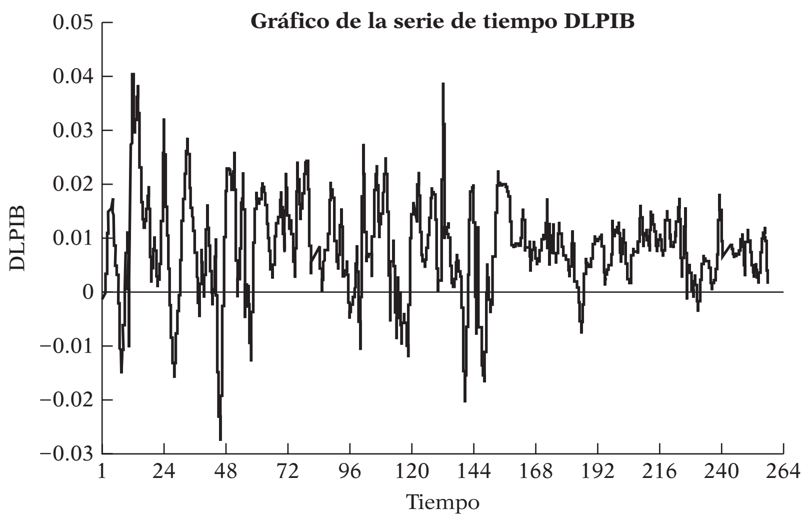
Al reconsiderar la serie de tiempo LPIB de Estados Unidos, ya vimos que tiene raíz unitaria. Ahora veremos lo que sucede si se toman las primeras diferencias de la serie LPIB.

Sea $\Delta \text{LPIB}_t = (\text{LPIB}_t - \text{LPIB}_{t-1})$. Por conveniencia, sea $D_t = \Delta \text{LPIB}_t$. Ahora considere la siguiente regresión:

$$\widehat{\Delta D_t} = 0.00557 - 0.6711 D_{t-1}$$
$$t = (7.1407) \quad (-11.0204) \tag{21.10.1}$$
$$R^2 = 0.3360 \qquad d = 2.0542$$

El valor crítico $\tau$ a 1% para la DF es $-3.4574$. Como la $\tau$ calculada ($= t$) de $-11.0204$ es más negativa que el valor crítico, concluimos que la serie LPIB en primeras diferencias es estacionaria; o sea, es $I(0)$, como se muestra en la figura 21.9. Si comparamos esta figura con la 21.1, observará las evidentes diferencias entre ambas.

**FIGURA 21.9**
Primeras diferencias de los logaritmos del PIB de Estados Unidos, 1947-2007 (trimestral).




[38] Un estudio de estas pruebas se encuentra en Maddala *et al., op. cit.*, cap. 4.

[39] Si una serie de tiempo es $I(2)$, contendrá dos raíces unitarias, en cuyo caso tendremos que diferenciar dos veces. Si es $I(d)$, debe diferenciarse $d$ veces, donde $d$ es cualquier entero.

## Procesos estacionarios en tendencia

Como vimos en la figura 21.5, un PET es estacionario alrededor de la línea de tendencia. Por tanto, la manera más sencilla de convertir en estacionaria una serie de tiempo es hacer la regresión de ella sobre el tiempo y los residuos de tal regresión serán estacionarios. En otras palabras, realizamos la siguiente regresión:

$$Y_t = \beta_1 + \beta_2 t + u_t \tag{21.10.2}$$

donde $Y_t$ es la serie de tiempo estudiada y $t$ es la variable de tendencia medida de manera cronológica.

Ahora bien,

$$\hat{u}_t = (Y_t - \hat{\beta}_1 - \hat{\beta}_2 t) \tag{21.10.3}$$

será estacionaria. A $\hat{u}_t$ se le conoce como **serie de tiempo sin tendencia**.

Es importante notar que tal vez la tendencia sea no lineal. Por ejemplo, puede ser

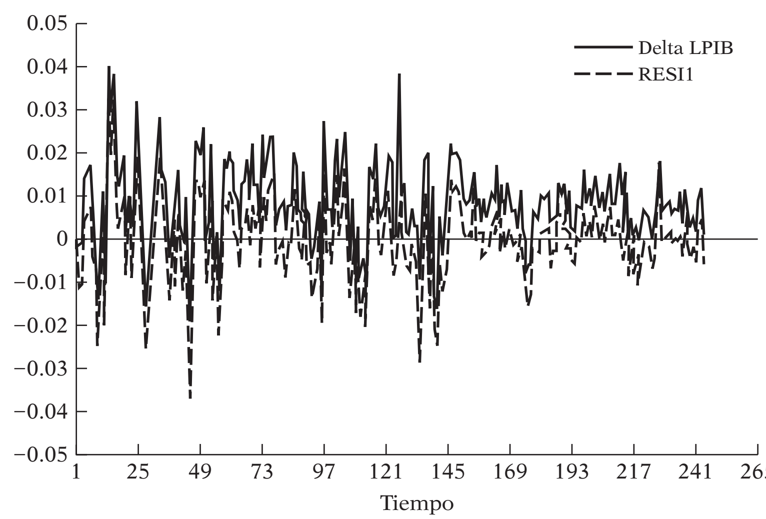
$$Y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + u_t \tag{21.10.4}$$

que es una serie con tendencia cuadrática. De ser así, los residuos de (21.10.4) serán ahora una serie (cuadrática) de tiempo sin tendencia.

Debe señalarse que si una serie de tiempo es PED pero se trata como si fuera PET, esto se conoce como **hipodiferenciación**. Por otra parte, si una serie de tiempo es PET pero se le trata como PED, se conoce como **hiperdiferenciación**. Las consecuencias de estos errores de especificación pueden ser graves, según la manera en que se manejen las propiedades de correlación de los términos de error resultantes.[40]

Para ver qué sucede si se confunde una serie PET con una serie PED o viceversa, la figura 21.10 muestra las primeras diferencias de LPIB y los residuos del LPIB estimado a partir de la regresión PET (21.10.2):

**FIGURA 21.10** Primeras diferencias (delta LPIB) y desviaciones de la tendencia (RESI1) para el logaritmo del PIB, 1947-2007 (trimestral).




[40] Para un análisis detallado de esto, véase Maddala *et al., op. cit.*, sección 2.7.

Un vistazo a esta figura revela que las primeras diferencias del logaritmo del PIB real son estacionarias (como lo confirma la regresión [21.10.1]), pero los residuos de la línea de tendencia (RESI1) no.

En resumen, "... es muy importante aplicar el tipo correcto de transformación de estacionariedad a los datos si no son ya estacionarios. La mayoría de los mercados financieros generan datos sobre precios, tasas o rendimientos que son no estacionarios debido a una tendencia estocástica más que determinista. Rara vez es apropiado suprimir la tendencia de los datos ajustando una línea de tendencia y tomando desviaciones. En cambio, para suprimir la tendencia de los datos es preciso tomar las primeras diferencias, por lo general el logaritmo del precio o las tasas, porque entonces los datos estacionarios transformados corresponderán a los rendimientos del mercado".[41]

## 21.11 Cointegración: regresión de una serie de tiempo con raíz unitaria sobre otra serie de tiempo con raíz unitaria

Ya advertimos que la regresión de una serie de tiempo no estacionaria sobre otra no estacionaria puede causar una regresión espuria. Suponga que consideramos las series de tiempo LGCP y LIDP presentadas en la sección 21.1 (consulte los datos reales en el sitio Web del libro). Si somete estas series de manera individual a un análisis de raíz unitaria encontrará que ambas son $I(1)$; es decir, contienen una tendencia estocástica. Es muy posible que las dos series *compartan la misma tendencia común,* por lo que la regresión de una sobre la otra no será necesariamente espuria.

Para ser específicos, usaremos los datos de las series de tiempo económicas de Estados Unidos (véase la sección 21.1 y el sitio Web del libro) y ejecutaremos la siguiente regresión de LGCP sobre LIPD:

$$\text{LGCP}_t = \beta_1 + \beta_2 \text{LIDP}_t + u_t \qquad \textbf{(21.11.1)}$$

donde L significa logaritmo. $\beta_2$ es la elasticidad del gasto de consumo personal real respecto del ingreso personal disponible real. Para efectos ilustrativos, le denominaremos *elasticidad del consumo.* Esto se expresa como:

$$u_t = \text{LGCP}_t - \beta_1 - \beta_2 \text{LIDP}_t \qquad \textbf{(21.11.2)}$$

Suponga que ahora sometemos $u_t$ a un análisis de raíz unitaria y descubrimos que es estacionaria, es decir, $I(0)$. Ésta es una situación interesante, pues $\text{LGCP}_t$ y $\text{LIDP}_t$ son individualmente $I(1)$, es decir, tienen tendencias estocásticas, y su combinación lineal (21.11.2) es $I(0)$. Se puede decir que la combinación lineal cancela las tendencias estocásticas de las dos series. Si consideramos el consumo y el ingreso como dos variables $I(1)$, el ahorro (definido como ingreso menos consumo) puede ser $I(0)$. Como resultado, una regresión del consumo sobre el ingreso, como en (21.11.1), puede ser significativa (es decir, no espuria). En este caso decimos que las dos variables están **cointegradas**. En términos económicos, dos variables serán cointegradas si existe una relación de largo plazo, o de equilibrio, entre ambas. La teoría económica a menudo se expresa en términos de equilibrio, como la teoría monetaria cuantitativa de Fisher o la teoría de la paridad del poder adquisitivo (PPA), por mencionar algunas.

En resumen, en tanto se verifique que los residuos de las regresiones como (21.11.1) son $I(0)$ o estacionarios, la metodología tradicional de regresión (inclusive las pruebas $t$ y $F$) aprendida hasta ahora es aplicable a las series de tiempo (no estacionarias). La contribución valiosa de los conceptos de raíz unitaria, cointegración, etc., es que obligan a determinar si los residuos de la regresión son estacionarios. Como observa Granger: "Una prueba para la cointegración puede considerarse como una preprueba para evitar las situaciones de regresiones espurias".[42]

En el lenguaje de la teoría de la cointegración, una regresión como (21.11.1) se conoce como **regresión cointegrante**, y el parámetro de pendiente $\beta_2$ como **parámetro cointegrante**. El con-

[41] Carol Alexander, *op. cit.,* p. 324.

[42] C.W.J. Granger, "Developments in the Study of Co-Integrated Economic Variables", en *Oxford Bulletin of Economics and Statistics*, vol. 48, 1986, p. 226.

cepto de cointegración puede extenderse a un modelo de regresión que contenga $k$ regresoras, en cuyo caso se tendrán $k$ parámetros cointegrantes.

## Prueba de cointegración

En las publicaciones especializadas se han propuesto varios métodos para probar la cointegración. Aquí consideraremos un método relativamente sencillo: la prueba de raíz unitaria DF o DFA sobre los residuos estimados a partir de la regresión cointegrante.[43]

*Prueba de Engle-Granger (EG) o prueba de Engle-Granger aumentada (EGA)*

Ya sabemos cómo aplicar las pruebas de raíz unitaria DF o DFA. Sólo requerimos estimar una regresión como (21.11.1), obtener los residuos y utilizar la prueba DF o DFA.[44] Sin embargo, debe tomarse una precaución. Como la $u_t$ estimada se basa en el parámetro de cointegración *estimado* $\beta_2$, los valores críticos de significancia DF y DFA no son del todo apropiados. Engle y Granger calcularon estos valores, los cuales se encuentran en las referencias.[45] Por consiguiente, en el contexto actual, las pruebas DF y DFA se conocen como la **prueba de Engle-Granger (EG)** y la **prueba de Engle-Granger aumentada (EGA).** Sin embargo, varios paquetes de software reportan actualmente estos valores críticos junto con otros resultados.

Ilustraremos estas pruebas. Con los datos introducidos en la sección 21.1 y publicados en el sitio Web del libro, primero realizamos la regresión de LGCPC sobre LIPDC y obtuvimos la siguiente regresión:

$$\begin{aligned} \widehat{\text{LGCP}}_t &= \quad -0.1942 \quad + \quad 1.0114\text{LIDP}_t \\ t &= \quad (-8.2328) \qquad (348.5429) \\ & \qquad\qquad R^2 = 0.9980 \qquad d = 0.1558 \end{aligned} \tag{21.11.3}$$

Como LGCP y LIDP son no estacionarios en lo individual, existe la posibilidad de que esta regresión sea espuria. Pero cuando llevamos a cabo una prueba de raíz unitaria sobre los residuos obtenidos en (21.11.3), resultó lo siguiente:

$$\begin{aligned} \widehat{\Delta \hat{u}_t} &= -0.0764\hat{u}_{t-1} \\ t &= (-3.0458) \\ & \qquad R^2 = 0.0369 \qquad d = 2.5389 \end{aligned} \tag{21.11.4}$$

Los valores críticos asintóticos Engle-Granger a 5% y 10% son de alrededor de −3.34 y −3.04, respectivamente. Por tanto, los residuos de la regresión son no estacionarios en el nivel de 5%. Sería difícil aceptar esta causa, pues la teoría económica indica que debe haber una relación estable entre GCP e IPD.

Volveremos a estimar la ecuación (21.11.3) con la variable de tendencia y luego veremos si los residuos de esta ecuación son estacionarios. Primero presentaremos los resultados y después analizaremos lo que ocurre.

$$\begin{aligned} \widehat{\text{LGCP}}_t &= \quad 2.8130 + \quad 0.0037_t \; + \quad 0.5844\text{LIPD}_t \\ t &= (21.3491) \quad (22.9394) \quad (31.2754) \\ & \qquad\qquad\qquad R^2 = 0.9994 \qquad d = 0.2956 \end{aligned} \tag{21.11.3a}$$

[43] Esta diferencia existe entre pruebas de raíces unitarias y pruebas de cointegración. Como señalan David A. Dickey, Dennis W. Jansen y Daniel I. Thornton: "Las pruebas para raíces unitarias se realizan sobre series de tiempo univariadas [es decir, singulares]. En contraste, la cointegración trata con la relación entre un grupo de variables, en donde cada una (incondicionalmente) tiene una raíz unitaria". Véase su artículo "A Primer on Cointegration with an Application to Money and Income", en *Economic Review*, Federal Reserve Bank of St. Louis, marzo-abril de 1991, p. 59. Como el nombre lo indica, es una introducción excelente para la prueba de cointegración.

[44] Si GCP e IPD no están cointegrados, las combinaciones lineales que de ellos se hagan no serán estacionarias y, por consiguiente, los residuos $u_t$ tampoco lo serán.

[45] R.F. Engle y C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation and Testing", en *Econometrica*, vol. 55, 1987, pp. 251-276.

Para ver si los residuos de esta regresión son estacionarios, obtuvimos los siguientes resultados (compare con la ecuación [21.11.4]):

$$\widehat{\Delta u_t} = -0.1498\hat{u}_{t-1}$$
$$t = (-4.4545) \qquad \textbf{(21.11.4}a\textbf{)}$$
$$R^2 = 0.0758 \qquad d = 2.3931$$

*Nota:* $\hat{u}_t$ es el residuo de la ecuación (21.11.3*a*).

La prueba DF muestra ahora que estos residuos son estacionarios. Aunque se use la prueba DFA con varios rezagos, los residuos siguen siendo estacionarios.

¿Qué sucede aquí? Aunque los residuos de la regresión (21.11.4*a*) son estacionarios, es decir, $I(0)$, son estacionarios alrededor de una tendencia de tiempo determinista, que es lineal. Es decir, los residuos son $I(0)$ más una tendencia lineal. Como ya señalamos, una serie de tiempo puede contener tanto una tendencia determinista como una estocástica.

Antes de continuar, debe señalarse que estos datos de series de tiempo abarcan un periodo largo (61 años). Es muy posible que, a causa de cambios estructurales en la economía de Estados Unidos durante este periodo, los resultados y conclusiones difieran. En el ejercicio 21.28 se le pedirá comprobar esta posibilidad.

## Cointegración y mecanismo de corrección de errores (MCE)

Acabamos de demostrar que, teniendo en cuenta la tendencia (lineal), LGCP y LIPD están cointegradas, es decir, hay una relación de equilibrio de largo plazo entre las dos. Desde luego, en el corto plazo puede haber desequilibrio. En consecuencia, podemos tratar el término de error en la siguiente ecuación como el "error de equilibrio". Además, con este término de error podemos relacionar el comportamiento de corto plazo del GCP con su valor de largo plazo:

$$u_t = \text{LGCP}_t - \beta_1 - \beta_2\text{LIPD} - \beta_3 t \qquad \textbf{(21.11.5)}$$

El **mecanismo de corrección de errores (MCE)**, utilizado por primera vez por Sargan[46] y popularizado más tarde por Engle y Granger, corrige el desequilibrio. Un importante teorema, conocido como **teorema de representación de Granger**, afirma que si dos variables $Y$ y $X$ están cointegradas, la relación entre las dos se expresa como MCE. Para ver lo que esto significa, revertiremos el ejemplo de GCP e IPD. Ahora considere el siguiente modelo:

$$\Delta\text{LGCP}_t = \alpha_0 + \alpha_1\Delta\text{LIPD}_t + \alpha_2\text{u}_{t-1} + \varepsilon_t \qquad \textbf{(21.11.6)}$$

donde $\varepsilon_t$ es un término de error de ruido blanco y $u_{t-1}$ es el valor rezagado del término de error de la ecuación (21.11.5).

La ecuación MCE (21.11.5) establece que ΔLGCP depende de ΔLIPD y también del término de error de equilibrio.[47] Si este último es diferente de cero, el modelo no está en equilibrio. Suponga que ΔLIPD es cero y que $u_{t-1}$ es positiva. Esto significa que $\text{LGCP}_{t-1}$ es demasiado alto para estar en equilibrio, es decir, $\text{LGCP}_{t-1}$ está por encima de su valor de equilibrio $(\alpha_0 + \alpha_1\text{LIDP}_{t-1})$. Como se espera que $\alpha_2$ sea negativa, el término $\alpha_2 u_{t-1}$ es negativo y, por tanto, $\Delta\text{LGCP}_t$ será negativo para restablecer el equilibrio. Es decir, si $\text{LGCP}_t$ está por arriba de su valor de equilibrio, comenzará a disminuir en el siguiente periodo a fin de corregir el error de equilibrio; de ahí el nombre de MCE. De igual manera, si $u_{t-1}$ es negativa (es decir, LGCP está por debajo de su valor de equilibrio), $\alpha_2 u_{t-1}$ será positivo, lo cual provocará que $\Delta\text{LGCP}_t$ sea positivo, lo que provocará que $\text{LGCP}_t$ se incremente en el periodo $t$. Por tanto, el valor absoluto de $\alpha_2$ determina la rapidez con que se restablecerá el equilibrio. En la práctica, $u_{t-1}$ se estima

[46] J.D. Sargan, "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", publicado en K.F. Wallis y D.F. Hendry (eds.), *Quantitative Economics and Econometric Analysis*, Basil Blackwell, Oxford, Inglaterra, 1984.

[47] El siguiente análisis se basó en Gary Koop, *op. cit.*, pp. 159-160, y Kerry Peterson, *op. cit.*, sección 8.5.

por $\hat{u}_{t-1} = (\text{LGCP}_t - \hat{\beta}_1 - \hat{\beta}_2\text{LIPD} - \hat{\beta}_3 t)$. Tenga en cuenta que se espera que el coeficiente de corrección del error $\alpha_2$ sea negativo (¿por qué?).

De regreso al ejemplo ilustrativo, la contraparte empírica de (21.11.6) es:

$$\widehat{\Delta\text{LGCP}}_t = 0.0061 + 0.2967\Delta\text{LIDP}_t - 0.1223\hat{u}_{t-1}$$
$$t = (9.6753) \quad (6.2282) \quad (-3.8461) \tag{21.11.7}$$
$$R^2 = 0.1658 \qquad d = 2.1496$$

Estadísticamente, el término MCE es significativo, lo que indica que el GCP se ajusta al IPD con un rezago; sólo alrededor de 12% de la discrepancia entre el GCP de largo y corto plazos se corrige dentro de un trimestre.

En la regresión (21.11.7) observamos que la elasticidad del consumo de corto plazo es de casi 0.29. La elasticidad de largo plazo es de casi 0.58, lo cual se observa en la ecuación (21.11.3*a*).

Antes de concluir esta sección, es importante recordar la recomendación de S. G. Hall:

> Mientras que el concepto de cointegración es sin duda un fundamento teórico importante del modelo de corrección de errores, hay aún diversos problemas en torno a su aplicación práctica; los valores críticos y el desempeño en muestras pequeñas de muchas de las pruebas son desconocidos para un amplio rango de modelos; la inspección bien informada del correlograma puede ser aún una herramienta importante.[48]

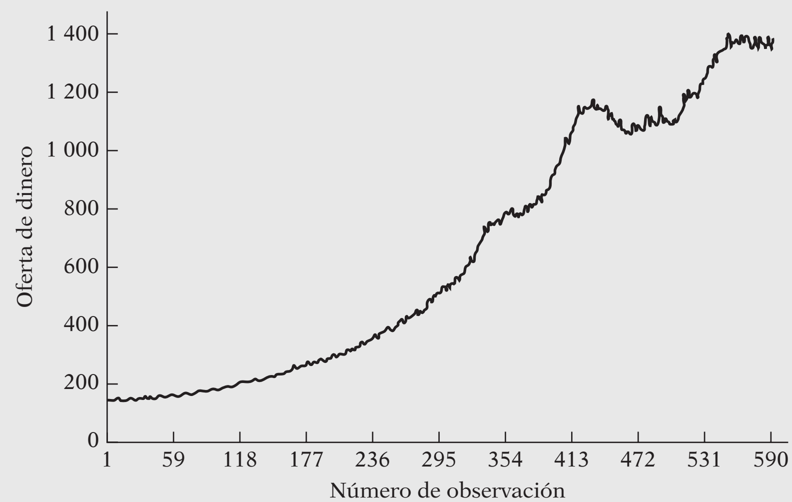
## 21.12 Algunas aplicaciones económicas

Concluimos este capítulo con el examen de algunos ejemplos concretos.

**EJEMPLO 21.1**
*Oferta mensual de dinero M1 en Estados Unidos, de enero de 1959 a 1 de marzo de 2008*

La figura 21.11 muestra la oferta de dinero M1 en Estados Unidos de enero de 1959 a 1 de marzo de 2008. De lo que sabemos sobre la estacionariedad, parece que la serie de tiempo oferta de dinero M1 es no estacionaria, lo cual se confirma mediante un análisis de raíz unitaria.

**FIGURA 21.11** Oferta de dinero en Estados Unidos de enero de 1959 a marzo de 2008.




(*continúa*)

[48] S.G. Hall, "An Application of the Granger and Engle Two-Step Estimation Procedure to the United Kingdom Aggregate Wage Data", en *Oxford Bulletin of Economics and Statistics*, vol. 48, núm. 3, agosto de 1986, p. 238. Véase también John Y. Campbell y Pierre Perron, "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots", en NBER (National Bureau of Economic Research), *Macroeconomics Annual 1991*, pp. 141-219.

**EJEMPLO 21.1**
(*continuación*)

(*Nota:* Para ahorrar espacio, no se dan los datos reales, pero pueden obtenerse de la Federal Reserve Board o el Federal Reserve Bank of St. Louis.)

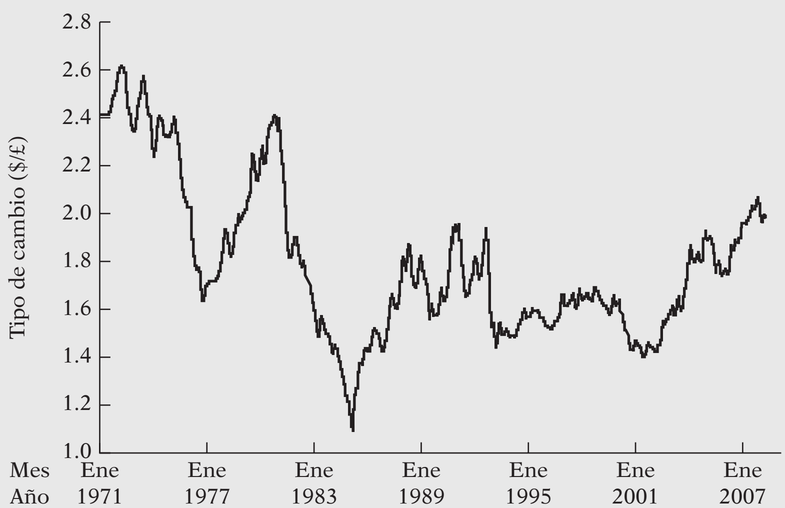
$$\Delta\hat{M}_t = -0.1347 + 0.0293t - 0.0102M_{t-1}$$
$$t = (-0.14) \quad (2.62) \quad (-2.30) \qquad \textbf{(21.12.1)}$$
$$R^2 = 0.0130 \qquad d = 2.2325$$

Los valores críticos $\tau$ a 1%, 5% y 10% son −3.9811, −3.4210 y −3.1329, respectivamente. Como el valor $t$ de −2.30 es menos negativo que cualquiera de estos valores críticos, la conclusión es que la serie de tiempo M1 es no estacionaria; o sea, contiene una raíz unitaria o es $I(1)$. Aunque se introdujeron diversos valores rezagados de $\Delta M_t$ (al estilo DFA), la conclusión no varió. Por otra parte, descubrimos que las primeras diferencias de la oferta de dinero M1 eran estacionarias (verifique esto).

**EJEMPLO 21.2**
*Tipo de cambio EUA/RU: Enero de 1971 a abril de 2008*

La figura 21.12 proporciona la gráfica del tipo de cambio ($/£) de enero de 1971 a abril de 2008 para un total de 286 observaciones. A estas alturas, el lector debe reconocer que esta serie de tiempo es no estacionaria. Al efectuar las pruebas de raíz unitaria, obtuvimos los siguientes estadísticos $\tau$: −0.82 (sin intercepto ni tendencia), −1.96 (intercepto) y −1.33 (con intercepto y tendencia). Cada uno de ellos, en valor absoluto, fue menor que su valor crítico $\tau$ tomado de las tablas DF apropiadas, por lo cual se confirma la impresión gráfica de que la serie de tiempo del tipo de cambio EUA/RU es no estacionaria.

**FIGURA 21.12**
Tipo de cambio Estados Unidos/RU: enero de 1971 a abril de 2008.



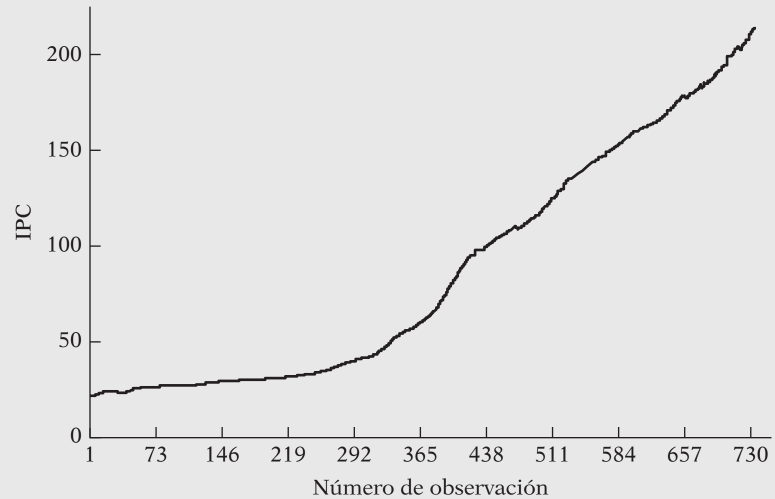

**EJEMPLO 21.3**
*Índice de precios al consumidor (IPC) en Estados Unidos de enero de 1947 a marzo de 2008*

La figura 21.13 muestra el IPC en Estados Unidos de enero de 1947 a marzo de 2008, para un total de 733 observaciones. La serie IPC, al igual que la serie M1 ya considerada, muestra una tendencia ascendente sostenida. El ejercicio de raíz unitaria proporcionó los siguientes resultados:

$$\widehat{\Delta \text{CPI}}_t = -0.01082 + 0.00068t - 0.00096\text{CPI}_{t-1} + 0.40669\Delta\text{CPI}_{t-1}$$
$$t = (-0.54) \quad (4.27) \quad (-1.77) \quad (12.03) \qquad \textbf{(21.12.2)}$$
$$R^2 = 0.3570 \qquad d = 1.9295$$

**EJEMPLO 21.3** (*continuación*)

**FIGURA 21.13** IPC en Estados Unidos, enero de 1947 a marzo de 2008.




El valor $t$ ($= \tau$) del $IPC_{t-1}$ es $-1.77$. El valor crítico a 10% es $-3.1317$. Como, en términos absolutos, la $\tau$ calculada es menor que la $\tau$ crítica, la conclusión es que el IPC no es una serie de tiempo estacionaria. Podemos caracterizar lo anterior como una tendencia estocástica (¿por qué?). Sin embargo, si tomamos las primeras diferencias de la serie IPC, descubrirá que son estacionarias. Por tanto, el IPC es una serie de tiempo estacionaria en diferencias (ED).

**EJEMPLO 21.4** *¿Están cointegradas las tasas de los pagarés del Tesoro a tres y seis meses?*

La figura 21.14 presenta la gráfica de las tasas de los pagarés del Tesoro de Estados Unidos (con vencimiento constante) a tres y seis meses, de enero de 1982 a marzo de 2008, para un total de 315 observaciones. ¿Muestra la gráfica que las dos tasas están cointegradas? Es decir, ¿existe una relación de equilibrio entre ambas? Con base en la teoría financiera, se podría esperar que así sucediera; de otra manera, los árbitros aprovecharían cualquier discrepancia entre las tasas de corto y largo plazos. En primer lugar, veamos si las dos series de tiempo son estacionarias.

**FIGURA 21.14** Tasas de los pagarés del Tesoro de Estados Unidos a tres y seis meses (con vencimiento constante).

(*continúa*)

**EJEMPLO 21.4** (*continuación*)

Con base en el modelo de caminata puramente aleatoria (es decir, sin intercepto ni tendencia), ambas tasas fueron estacionarias. Después de incluir el intercepto y la tendencia, así como una diferencia rezagada, el resultado señaló que las dos tasas pueden ser estacionarias en tendencia; el coeficiente de tendencia en ambos casos fue negativo y significativo en un nivel de 7%. Así, según los resultados que se acepten, las dos tasas son estacionarias o estacionarias en tendencia.

Al hacer la regresión de la tasa de los pagarés del Tesoro a seis meses (TB6) sobre la de tres meses, obtuvimos la siguiente regresión:

$$\begin{aligned}\widehat{\text{TB6}}_t &= \ 0.0842 \ + 1.0078\text{TB3}_t \\ t &= (3.65) \quad (252.39) \\ & \qquad R^2 = 0.995 \qquad d = 0.4035\end{aligned} \tag{21.12.3}$$

Al aplicar la prueba de raíz unitaria a los residuos de la regresión anterior descubrimos que los residuos eran estacionarios, lo cual indica que las tasas de interés de los pagarés a tres y seis meses estaban cointegradas. Con este conocimiento obtuvimos el siguiente modelo de corrección del error (MCE):

$$\begin{aligned}\Delta\widehat{\text{TB6}}_t &= \ -0.0047 + \ 0.8992\Delta\text{TB3}_t - \ 0.1855\hat{u}_{t-1} \\ t &= (-0.82) \qquad (47.77) \qquad\qquad (-5.69) \\ & \qquad\qquad R^2 = 0.880 \qquad d = 1.5376\end{aligned} \tag{21.12.4}$$

donde $\hat{u}_{t-1}$ es el valor rezagado del término de corrección del error para el periodo anterior. Como muestran estos resultados, 0.19 de la discrepancia en las dos tasas de interés del mes anterior se elimina al siguiente mes.[49] Además, los cambios de corto plazo en las tasas de interés de los pagarés del Tesoro a tres meses se reflejan de inmediato en la tasa de interés de los pagarés del Tesoro a seis meses, pues el coeficiente de la pendiente entre las dos es 0.8992. Esto no debe sorprender en vista de la eficiencia de los mercados de dinero de Estados Unidos.

## Resumen y conclusiones

1. El análisis de regresión basado en información de series de tiempo supone implícitamente que las series de tiempo en las cuales se basa son estacionarias. Las pruebas clásicas $t$ y $F$, entre otras, se basan en este supuesto.
2. En la práctica, la mayoría de las series de tiempo económicas son no estacionarias.
3. Decimos que un proceso estocástico es **estacionario débil** si su media, varianza y autocovarianzas son constantes en el tiempo (es decir, son invariantes en el tiempo).
4. En un nivel informal, la estacionariedad débil se prueba mediante el **correlograma** de una serie de tiempo, que es una gráfica de la autocorrelación en diferentes rezagos. Para una serie de tiempo estacionaria, el correlograma se desvanece rápidamente, mientras que para las series no estacionarias, lo hace de manera gradual. Cuando una serie es puramente aleatoria, las autocorrelaciones en todos los rezagos 1 y superiores son cero.
5. En un nivel formal, la estacionariedad se verifica averiguando si la serie de tiempo contiene una raíz unitaria. Las pruebas de **Dickey-Fuller (DF)** y **Dickey-Fuller Aumentada (DFA)** sirven para este propósito.
6. Una serie de tiempo económica puede ser **estacionaria en tendencia (ET)** o **estacionaria en diferencia (ED)**. Una serie de tiempo ET tiene tendencia determinista, mientras que una serie de tiempo ED tiene tendencia variable o estocástica. La práctica común de incluir la

[49] Como ambas tasas de interés de los pagarés del Tesoro se expresan en forma porcentual, esto indicaría que si la tasa de interés de los pagarés del Tesoro a seis meses fuera mayor que la tasa a tres meses, en una cantidad mayor que la esperada *a priori* en el último mes, el siguiente mes ésta se reduciría en 0.19 puntos porcentuales a fin de restablecer la relación de largo plazo entre las dos tasas de interés. Para conocer más sobre la teoría en la que se basa la relación entre las tasas de interés de corto y largo plazos, consulte cualquier libro de texto sobre banca o dinero y lea sobre la estructura de los plazos de las tasas de interés.

variable de tiempo o de tendencia en un modelo de regresión para eliminar la influencia de la tendencia en los datos sólo se justifica para series de tiempo ET. Las pruebas DF y DFA se aplican para determinar si una serie de tiempo es ET o ED.

7. La regresión de una variable de serie de tiempo sobre una o más variables de series de tiempo a menudo puede dar resultados sin sentido o espurios. Este fenómeno se conoce como **regresión espuria**. Una forma de evitarla es establecer si las series de tiempo están cointegradas.
8. **Cointegración** significa que, a pesar de no ser estacionarias en un nivel individual, una combinación lineal de dos o más series de tiempo puede ser estacionaria. Las pruebas **Engle-Granger (EG)** y **Engle-Granger aumentada (EGA)** sirven para averiguar si dos o más series de tiempo están cointegradas.
9. La cointegración de dos (o más) series de tiempo indica que existe una relación de largo plazo, o de equilibrio, entre ellas.
10. El **mecanismo de corrección de errores (MCE)** de Engle y Granger sirve para conciliar el comportamiento de corto plazo de una variable económica con su comportamiento de largo plazo.
11. El campo de la econometría de series de tiempo ha evolucionado. Los resultados y pruebas establecidas en algunos casos son tentativos y queda aún mucho trabajo pendiente. Una pregunta importante, aún pendiente, es por qué algunas series de tiempo económicas son estacionarias y otras no lo son.

# EJERCICIOS

## Preguntas

21.1. ¿Qué significa estacionariedad débil?

21.2. ¿Qué significa serie de tiempo integrada?

21.3. ¿Cuál es el significado de raíz unitaria?

21.4. Si una serie de tiempo es $I(3)$, ¿cuántas veces debe diferenciarse para hacerla estacionaria?

21.5. ¿Qué son las pruebas Dickey-Fuller (DF) y DF aumentada?

21.6. ¿Qué son las pruebas Engle-Granger (EG) y EG aumentada?

21.7. ¿Cuál es el significado de cointegración?

21.8. ¿Cuál es la diferencia, si acaso, entre pruebas de raíz unitaria y pruebas de cointegración?

21.9. ¿Qué es la regresión espuria?

21.10. ¿Cuál es la conexión entre cointegración y regresión espuria?

21.11. ¿Cuál es la diferencia entre una tendencia determinista y una tendencia estocástica?

21.12. ¿Qué significa proceso estacionario en tendencia (PET) y proceso estacionario en diferencias (PED)?

21.13. ¿Qué es una caminata aleatoria (modelo)?

21.14. "Para un proceso estocástico de caminata aleatoria, la varianza es infinita." ¿Está de acuerdo? ¿Por qué?

21.15. ¿Qué es el mecanismo de corrección de errores (MCE)? ¿Cuál es su relación con la cointegración?

## Ejercicios empíricos

21.16. Con los datos de series de tiempo económicas de Estados Unidos que se presentan en el sitio Web del libro, obtenga los correlogramas muestrales hasta 36 rezagos para las series de tiempo LGCP, LIPD, LUE (utilidades empresariales) y LDIVIDENDOS. ¿Qué patrón general observa? Por intuición, ¿cuáles de estas series parecen estacionarias?

21.17. Para cada una de las series de tiempo del ejercicio 21.16, utilice la prueba DF para determinar si estas series contienen una raíz unitaria. Si existe una raíz unitaria, ¿cómo caracteriza esa serie de tiempo?

21.18. Continúe con el ejercicio 21.17. ¿Cómo determina si una prueba DFA es más apropiada que una prueba DF?

21.19. Considere las series de tiempo de dividendos y utilidades contenidas en los datos sobre la economía de Estados Unidos publicados en el sitio Web del libro. Como los dividendos dependen de las utilidades, considere el siguiente modelo simple:

$$\text{LDIVIDENDOS}_t = \beta_1 + \beta_2\text{LUE} + u_t$$

*a*) ¿Esperaría que esta regresión sufra del fenómeno de regresión espuria? ¿Por qué?

*b*) ¿Están cointegradas las series de tiempo de los logaritmos de los dividendos y las utilidades? ¿Cómo probar esto explícitamente? Si después de la prueba encuentra que están cointegradas, ¿cambiaría la respuesta en *a*)?

*c*) Con el mecanismo de corrección de errores (MCE) estudie el comportamiento de corto y largo plazos de los dividendos en relación con las utilidades.

*d*) Si examina las series de LDIVIDENDOS y LUE individualmente, ¿presentan tendencias estocásticas o deterministas? ¿Qué pruebas utiliza?

**e*) Suponga que LDIVIDENDOS y LUE están cointegradas. Entonces, en lugar de efectuar la regresión de los dividendos sobre las utilidades, hace la regresión de las utilidades sobre los dividendos. ¿Es válida tal regresión?

21.20. Obtenga las primeras diferencias de las series de tiempo contenidas en los datos sobre la economía de Estados Unidos que se presentan en el sitio Web del libro y grafíquelas. Obtenga también un correlograma de cada serie de tiempo hasta 36 rezagos. ¿Qué le llama la atención sobre estos correlogramas?

21.21. En lugar de efectuar la regresión de LDIVIDENDOS sobre LUE en la forma de nivel, suponga que efectúa la regresión de las primeras diferencias de LDIVIDENDOS sobre las primeras diferencias de LUE. ¿Incluiría el intercepto en esta regresión? ¿Por qué? Muestre los cálculos.

21.22. Continúe con el ejercicio anterior. ¿Cómo probaría la presencia de estacionariedad en la regresión de primeras diferencias? En este ejemplo, ¿qué esperaría *a priori* y por qué? Muestre todos los cálculos.

21.23. Con base en el número de nuevas viviendas construidas en el Reino Unido ($X$) de 1948 a 1984, Terence Mills obtuvo los siguientes resultados de regresión:[†]

$$\begin{aligned}\widehat{\Delta X_t} &= \quad 31.03 \quad - \quad 0.188X_{t-1}\\ \text{se} &= \ (12.50) \qquad (0.080)\\ (t =)\tau & \qquad\qquad\quad (-2.35)\end{aligned}$$

*Nota:* El valor crítico $\tau$ a 5% es $-2.95$ y el valor crítico $\tau$ a 10% es $-2.60$.

*a*) Con base en estos resultados, ¿la serie de tiempo de nuevas construcciones de viviendas es estacionaria o no estacionaria? Por otra parte, ¿hay una raíz unitaria en esta serie de tiempo? ¿Cómo sabe?

*b*) Si fuera a utilizar la prueba $t$ usual, ¿es el valor $t$ observado estadísticamente significativo? Con esta base, ¿habría concluido que esta serie de tiempo es estacionaria?

*c*) Ahora considere los siguientes resultados de regresión:

$$\begin{aligned}\widehat{\Delta^2 X_t} &= \ 4.76 \ - 1.39\Delta X_{t-1} + 0.313\Delta^2 X_{t-1}\\ \text{se} &= (5.06) \quad (0.236) \qquad\quad (0.163)\\ (t =)\tau & \qquad\quad (-5.89)\end{aligned}$$

* Opcional.

† Terence C. Mills, *op. cit.*, p. 127. La notación se alteró un poco.

donde $\Delta^2$ es el operador de segundas diferencias, es decir, primeras diferencias de las primeras diferencias. El valor estimado es ahora estadísticamente significativo. ¿Qué puede decir sobre la estacionariedad de la serie de tiempo en cuestión?

*Nota:* El propósito de la regresión anterior es determinar si hay **una segunda raíz unitaria** en la serie de tiempo.

21.24. Genere dos series de caminata aleatoria como se indica en (21.7.1) y (21.7.2) y haga la regresión de una sobre la otra. Repita este ejercicio pero ahora con sus primeras diferencias y verifique que en esta regresión el valor de $R^2$ sea casi cero y que la $d$ de Durbin-Watson sea casi 2.

21.25. Para mostrar que dos variables, cada una con tendencia determinista, pueden originar una regresión espuria, Charemza *et al.* obtuvieron la siguiente regresión con base en 30 observaciones:*

$$\hat{Y}_t = 5.92 + 0.030X_t$$
$$t = (9.9) \quad (21.2)$$
$$R^2 = 0.92 \qquad d = 0.06$$

donde $Y_1 = 1, Y_2 = 2, \ldots, Y_n = n$ y $X_1 = 1, X_2 = 4, \ldots, X_n = n^2$.

*a*) ¿Qué tipo de tendencia muestra $Y$? ¿y $X$?

*b*) Grafique las dos variables y la línea de regresión. ¿Qué conclusión general obtiene de esta gráfica?

21.26. De los datos correspondientes del primer trimestre de 1971 al cuarto de 1988 para Canadá se obtuvieron los siguientes resultados de la regresión:

1. $$\widehat{\ln \text{M1}}_t = -10.2571 + 1.5975 \ln \text{PIB}_t$$
$$t = (-12.9422) \quad (25.8865)$$
$$R^2 = 0.9463 \qquad d = 0.3254$$

2. $$\widehat{\Delta \ln \text{M1}}_t = 0.0095 + 0.5833 \Delta \ln \text{PIB}_t$$
$$t = (2.4957) \quad (1.8958)$$
$$R^2 = 0.0885 \qquad d = 1.7399$$

3. $$\Delta\hat{u}_t = -0.1958\hat{u}_{t-1}$$
$$(t = \tau) \; (-2.2521)$$
$$R^2 = 0.1118 \qquad d = 1.4767$$

donde M1 = la oferta de dinero M1, PIB = producto interno bruto, ambas medidas en miles de millones de dólares canadienses, ln es el logaritmo natural y $\hat{u}_t$ representa los residuos estimados de la regresión 1.

*a*) Interprete las regresiones 1 y 2.

*b*) ¿Sospecha que la regresión 1 es espuria? ¿Por qué?

*c*) ¿La regresión 2 es espuria? ¿Cómo sabe?

*d*) De los resultados de la regresión 3, ¿cambiaría su conclusión de *b*)? ¿Por qué?

* Charemza *et al.*, *op. cit.*, p. 93.

*e*) Ahora considere la siguiente regresión:

$$\widehat{\Delta \ln \mathrm{M1}_t} = 0.0084 + 0.7340\Delta \ln \mathrm{PIB}_t - 0.0811\hat{u}_{t-1}$$
$$t = (2.0496) \quad (2.0636) \quad (-0.8537)$$
$$R^2 = 0.1066 \qquad d = 1.6697$$

¿Qué indica esta regresión? ¿Le ayuda a decidir si la regresión 1 es espuria o no lo es?

21.27. Las siguientes regresiones se basan en los datos del IPC de Estados Unidos de 1960 a 2007, para un total de 48 observaciones anuales:

1. $$\widehat{\Delta \mathrm{IPC}_t} = 0.0334\mathrm{IPC}_{t-1}$$
$$t = (12.37)$$
$$R^2 = 0.0703 \qquad d = 0.3663 \qquad \mathrm{SCR} = 206.65$$

2. $$\widehat{\Delta \mathrm{IPC}_t} = 1.8662 + 0.0192\mathrm{IPC}_{t-1}$$
$$t = (3.27) \quad (3.86)$$
$$R^2 = 0.249 \qquad d = 0.4462 \qquad \mathrm{SCR} = 166.921$$

3. $$\widehat{\Delta \mathrm{IPC}_t} = 1.1611 + 0.5344t - 0.1077\mathrm{IPC}_{t-1}$$
$$t = (2.37) \quad (4.80) \quad (-4.02)$$
$$R^2 = 0.507 \qquad d = 0.6071 \qquad \mathrm{SCR} = 109.608$$

donde SCR = suma de cuadrados residual.

*a*) Al examinar las regresiones anteriores, ¿qué puede decir respecto de la estacionariedad de la serie de tiempo IPC?

*b*) ¿Cómo escogería entre los tres modelos?

*c*) La ecuación (1) es la ecuación (3) menos el intercepto y la tendencia. ¿Con qué prueba decidiría si las restricciones implícitas del modelo 1 son válidas? (*Sugerencia:* Utilice las pruebas Dickey-Fuller $t$ y $F$. Use los valores aproximados dados en el **apéndice D**, tabla D.7.)

21.28. Como indicamos en el texto, puede haber varias rupturas estructurales en el conjunto de datos de series de tiempo económicas de Estados Unidos de la sección 21.1. Las variables dicótomas son una buena forma de incorporar estos cambios en los datos.

*a*) Con variables dicótomas para designar tres periodos diferentes relacionados con los embargos petroleros de 1973 y 1979, efectúe una regresión del logaritmo de gasto de consumo personal (LGCP) sobre el logaritmo del ingreso personal disponible (LIPD). ¿Hubo algún cambio en los resultados? ¿Cuál es ahora su decisión sobre la hipótesis de raíz unitaria?

*b*) Varios sitios Web presentan los ciclos económicos oficiales que pueden haber afectado los datos de series de tiempo económicas de Estados Unidos que analizamos en la sección 21.1. Consulte, por ejemplo, http://www.nber.org/cycles/cyclesmain.html. Con la información que ahí se encuentra, cree variables dicótomas que indiquen algunos de los principales ciclos y verifique los resultados de la regresión de LGCP sobre LIPD. ¿Hubo algún cambio?

# Capítulo 22

# Econometría de series de tiempo: pronósticos

En la **introducción** dijimos que los pronósticos son una parte importante del análisis econométrico, y para algunas personas constituye el área más importante. ¿Cómo se pronostican variables económicas como el PIB, inflación, tasas de cambio, precios de acciones, tasas de desempleo y miles de otras variables económicas? En este capítulo analizaremos dos métodos para pronosticar que han adquirido mucha popularidad: 1) el **autorregresivo integrado de promedios móviles (ARIMA)**, conocido como metodología de **Box-Jenkins**,[1] y 2) el de **vectores autorregresivos (VAR)**.

En este capítulo también estudiaremos los problemas especiales de pronosticar precios de los activos financieros, como los de las acciones y las tasas de cambio. Estos precios de los activos se caracterizan por un fenómeno conocido como **acumulación de volatilidad**, lo que significa que existen lapsos en los que muestran amplias variaciones durante prolongados periodos, seguidos por un intervalo de tranquilidad relativa. Basta observar el índice Dow Jones de los últimos tiempos. Los llamados **modelos con heteroscedasticidad condicional autorregresiva (ARCH)** o **modelos con heteroscedasticidad condicional autorregresiva generalizada (GARCH)** representan la mencionada acumulación de volatilidad.

El tema de los pronósticos económicos es amplio, por lo que se han escrito libros especializados sobre esta materia. El objetivo de este capítulo consiste en proporcionar al lector sólo una pequeña muestra de este tema. El lector interesado puede consultar la bibliografía para un estudio más profundo. Por fortuna, la mayoría del software moderno para la econometría contiene fáciles introducciones a las diversas técnicas que estudiamos aquí.

El vínculo entre este capítulo y el anterior reside en que los métodos de pronósticos analizados antes presuponen que las series de tiempo subyacentes son estacionarias o que pueden convertirse en estacionarias mediante transformaciones adecuadas. Conforme avancemos a lo largo de este capítulo veremos la utilización de diversos conceptos que presentamos en el capítulo anterior.

## 22.1 Enfoques de los pronósticos económicos

En términos generales, hay cinco enfoques de los pronósticos económicos basados en series de tiempo: 1) métodos de suavizamiento exponencial, 2) modelos de regresión uniecuacionales, 3) modelos de regresión de ecuaciones simultáneas, 4) modelos autorregresivos integrados de promedios móviles (ARIMA) y 5) modelos de vectores autorregresivos (VAR).

[1] G.P.E. Box y G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, edición revisada, Holden Day, San Francisco, 1978.

## Métodos de suavizamiento exponencial[2]

En esencia son métodos para ajustar una curva apropiada a datos históricos de una determinada serie de tiempo. Existen diversos tipos, como el de *suavizamiento exponencial simple*, el *método lineal de Holt* y el *método de Holt-Winters*, así como sus variaciones. Aunque todavía se emplean en varias áreas para pronósticos en los negocios y la economía, ahora se complementan (¿o sustituyen?) por los otros cuatro métodos mencionados. No los analizaremos en este capítulo, pues se desvían mucho de la materia.

## Modelos de regresión uniecuacionales

Dedicamos una parte sustancial de esta obra a los modelos de regresión uniecuacionales. Como ejemplo de un modelo de regresión uniecuacional, considere la función de demanda de automóviles. La teoría económica postula que la demanda de automóviles es función de sus precios, gasto en publicidad, ingreso del consumidor, tasa de interés (como medida del costo de endeudamiento) y otras variables relevantes (por ejemplo, tamaño de la familia o distancia al trabajo). A partir de las series de tiempo estimamos un modelo apropiado de demanda de automóviles (lineal, log-lineal o no lineal), el cual sirve para el pronóstico de la demanda en el futuro. Por supuesto, como vimos en el capítulo 5, los errores de pronóstico aumentan rápidamente si se va demasiado lejos hacia el futuro.

## Modelos de regresión de ecuaciones simultáneas[3]

En los capítulos 18, 19 y 20 consideramos los modelos de ecuaciones simultáneas. En su momento de apogeo, durante las décadas de 1960 y 1970, los modelos elaborados para describir la economía de Estados Unidos basados en ecuaciones simultáneas dominaron el campo de los pronósticos económicos. Pero más adelante, el encanto por ese tipo de pronósticos terminó debido a las crisis del precio del petróleo de 1973 y de 1979 y a la **crítica de Lucas**.[4] El argumento de esta crítica es que los parámetros estimados de un modelo econométrico dependen de la política prevaleciente en el momento en que se estima el modelo y cambian conforme lo hace la política. En resumen, los parámetros estimados no son invariantes ante cambios de política.

Por ejemplo, en octubre de 1979, el Banco de la Reserva Federal cambió su política monetaria en forma sustancial. En lugar de fijar metas de tasas de interés, anunció que en adelante supervisaría la tasa de crecimiento de la oferta monetaria. Ante un cambio tan relevante, un modelo econométrico estimado a partir de información pasada tendría poco valor predictivo con el nuevo régimen. En estos tiempos, el interés del Banco de la Reserva Federal pasó de controlar la oferta monetaria a controlar la tasa de interés de corto plazo (la tasa de los fondos federales).

## Modelos ARIMA

La publicación de G. P. E. Box y G. M. Jenkins *Time Series Analysis: Forecasting and Control, op. cit.,* marcó el comienzo de una nueva generación de herramientas de pronóstico. Popularmente conocida como metodología de Box-Jenkins (BJ), pero técnicamente conocida como metodología ARIMA, el interés de estos métodos de pronósticos no está en la construcción de modelos uniecuacionales o de ecuaciones simultáneas, sino en el análisis de las propiedades probabilísticas, o estocásticas, de las series de tiempo económicas por sí mismas según la filosofía

---

[2] Hay una exposición relativamente sencilla de estos métodos en Spyros Makridakis, Steven C. Wheelwright y Rob J. Hyndman, *Forecasting Methods and Applications*, 3a. ed., John Wiley & Sons, Nueva York, 1998.

[3] Para conocer un tratamiento didáctico acerca del uso de modelos de ecuaciones simultáneas en pronósticos, véase Robert S. Pindyck y Daniel L. Rubinfeld, *Econometric Models & Economic Forecasts*, McGraw-Hill, 4a. ed., Nueva York, 1998, parte III.

[4] Robert E. Lucas, "Econometric Policy Evaluation: A Critique", en Carnegie-Rochester Conference Series, *The Phillips Curve*, North-Holland, Ámsterdam, 1976, pp. 19-46. Este artículo, entre otros, hizo posible que Lucas obtuviera el Premio Nobel de Economía.

de *que los datos hablen por sí mismos*. A diferencia de los modelos de regresión, en los cuales $Y_t$ se explica por las $k$ regresoras $X_1, X_2, X_3, \ldots, X_k$, en los modelos de series de tiempo del tipo BJ, $Y_t$ se explica por valores pasados o rezagados de sí misma y por los términos de error estocásticos. Por esta razón, los modelos ARIMA reciben algunas veces el nombre de modelos *ateóricos* —porque no se derivan de teoría económica alguna—, y las teorías económicas a menudo son la base de los modelos de ecuaciones simultáneas.

A propósito, observe que en este capítulo la atención se centra en los modelos ARIMA *univariados*, es decir, en los modelos ARIMA que pertenecen a una sola serie de tiempo. No obstante, el análisis puede extenderse a modelos ARIMA multivariados.

### Modelos VAR

A primera vista, la metodología **VAR** se asemeja a los modelos de ecuaciones simultáneas, pues considera diversas variables endógenas de manera conjunta. Pero cada variable endógena se explica por sus valores rezagados, o pasados, y por los valores rezagados de todas las demás variables endógenas en el modelo; usualmente no hay variables exógenas en el modelo.

En lo restante de este capítulo analizaremos las bases de los enfoques de Box-Jenkins y VAR para los pronósticos económicos. El análisis es elemental y heurístico. Se aconseja al lector que desee continuar el estudio de este tema consultar las referencias.[5]

## 22.2 Creación de modelos AR, PM y ARIMA para series de tiempo

Para presentar diversas ideas, antiguas y nuevas, aquí trabajaremos con la serie de tiempo del PIB de Estados Unidos de la sección 21.1 (consulte los datos reales en el sitio Web del libro). La gráfica de dicha serie de tiempo ya se ilustró en las figuras 21.1 (logaritmo del PIB no diferenciado) y 21.9 (LPIB en primeras diferencias); recuerde que el LPIB en su forma de nivel es no estacionario, pero en su forma de (primeras) diferencias sí lo es.

Si una serie de tiempo es estacionaria, se puede modelar en diversas formas.

### Proceso autorregresivo (AR)

Sea $Y_t$ el logaritmo del PIB en el periodo $t$. Si se modela $Y_t$ como

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \qquad \textbf{(22.2.1)}$$

donde $\delta$ es la media de $Y$ y $u_i$ es un término de error aleatorio no correlacionado con media cero y varianza constante $\sigma^2$ (es decir, *ruido blanco*), se dice que $Y_t$ sigue un proceso estocástico **autorregresivo de primer orden**, o **AR(1)**, del cual ya hablamos en el capítulo 12. Aquí el valor de $Y$ en el tiempo $t$ depende de su valor en el periodo anterior y de un término aleatorio; los valores de $Y$ están expresados como desviaciones de su valor medio. En otras palabras, este modelo dice que el valor de pronóstico de $Y$ en el periodo $t$ es simplemente alguna proporción ($= \alpha_1$) de su valor en el periodo $(t - 1)$ más un "choque" o perturbación aleatoria en el tiempo $t$; de nuevo, los valores de $Y$ están expresados alrededor del valor de su media.

Pero si consideramos este modelo,

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_3(Y_{t-2} - \delta) + u_t \qquad \textbf{(22.2.2)}$$

[5] Véase Pindyck y Rubinfeld, *op. cit.*, parte 3; Alan Pankratz, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, Nueva York, 1991 (libro con aplicaciones); y Andrew Harvey, *The Econometric Analysis of Time Series*, The MIT Press, 2a. ed., Cambridge, Massachusetts, 1990 (libro relativamente avanzado). Un análisis completo pero fácil de entender se encuentra en Terence C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, Nueva York, 1990.

decimos que $Y_t$ sigue un **proceso autorregresivo de segundo orden** o **AR(2)**. Es decir, el valor de $Y$ en el tiempo $t$ depende de sus valores en los dos periodos anteriores, los valores de $Y$ expresados alrededor del valor de su media $\delta$.

En general, tenemos

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t \qquad \textbf{(22.2.3)}$$

en cuyo caso $Y_t$ es un **proceso autorregresivo de orden p**, o **AR(p)**.

Observe que en todos los modelos anteriores sólo se consideran los valores actuales y anteriores de $Y$; no hay otras regresoras. En este sentido, decimos que "los datos hablan por sí mismos". Son una clase de modelos de forma reducida a los cuales hicimos referencia en el análisis de los modelos de ecuaciones simultáneas.

## Proceso de medias móviles (MA)

El proceso AR recién expuesto no es el único mecanismo que pudo generar a $Y$. Suponga que hacemos un modelo de $Y$ de la siguiente manera:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \qquad \textbf{(22.2.4)}$$

donde $\mu$ es una constante y $u$, al igual que antes, es el término de error estocástico de ruido blanco. Aquí, $Y$ en el periodo $t$ es igual a una constante más un promedio móvil de los términos de error presente y pasado. Así, en este caso decimos que $Y$ sigue un proceso de **promedios móviles de primer orden**, o **MA(1)**.

Pero si $Y$ sigue la expresión

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} \qquad \textbf{(22.2.5)}$$

es un proceso **MA(2)**. En forma más general,

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \cdots + \beta_q u_{t-q} \qquad \textbf{(22.2.6)}$$

es un proceso **MA(*q*)**. En resumen, un proceso de promedios móviles es tan sólo una combinación lineal de términos de error de ruido blanco.

## Proceso autorregresivo y de promedios móviles (ARMA)

Desde luego, es muy probable que $Y$ tenga características de AR y de MA a la vez, y, por consiguiente, sea ARMA. Así, $Y_t$ sigue un proceso **ARMA(1, 1)** si se escribe como

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \qquad \textbf{(22.2.7)}$$

porque hay un término autorregresivo y uno de promedios móviles. En (22.2.7), $\theta$ representa un término constante.

En general, en un proceso **ARMA(*p*, *q*)**, habrá $p$ términos autorregresivos y $q$ términos de promedios móviles.

## Proceso autorregresivo integrado de promedios móviles (ARIMA)

Los modelos de series de tiempo analizados se basan en el supuesto de que las series de tiempo consideradas son (débilmente) estacionarias en el sentido definido en el capítulo 21. En pocas palabras, la media y la varianza de una serie de tiempo débilmente estacionaria son constantes y su covarianza es invariante en el tiempo. Pero sabemos que muchas series de tiempo económicas son no estacionarias, es decir, son integradas; por ejemplo, las series de tiempo económicas en la tabla 21.1 son integradas.

Sin embargo, vimos también en el capítulo 21 que si una serie de tiempo es integrada de orden 1 [es decir, si es $I(1)$], sus primeras diferencias son $I(0)$, es decir, estacionarias. En forma similar, si una serie de tiempo es $I(2)$, sus segundas diferencias son $I(0)$. En general, si una serie de tiempo es $I(d)$, después de diferenciarla $d$ veces se obtiene una serie $I(0)$.

Por consiguiente, si debemos diferenciar una serie de tiempo $d$ veces para hacerla estacionaria y luego aplicarle el modelo ARMA($p,q$), decimos que la serie de tiempo original es **ARIMA($p$, $d$, $q$)**, es decir, es una serie de tiempo **autorregresiva integrada de promedios móviles**, donde $p$ denota el número de términos autorregresivos, $d$ el número de veces que la serie debe diferenciarse para hacerse estacionaria y $q$ el número de términos de promedios móviles. Así, una serie de tiempo ARIMA(2, 1, 2) tiene que diferenciarse una vez ($d = 1$) antes de que se haga estacionaria, y la serie de tiempo estacionaria (en primeras diferencias) puede modelarse como un proceso ARMA(2, 2), es decir, tiene dos términos AR y dos términos MA. Desde luego, si $d = 0$ (es decir, si para empezar la serie es estacionaria), ARIMA($p$, $d = 0$, $q$) = ARMA($p$, $q$). Observe que un proceso ARIMA($p$, 0, 0) significa un proceso estacionario AR($p$) puro; un ARIMA (0, 0, $q$) significa un proceso estacionario MA($q$) puro. Con los valores de $p$, $d$ y $q$ sabemos de qué proceso se está haciendo el modelo.

El punto importante es que, para utilizar la metodología Box-Jenkins, debemos tener una serie de tiempo estacionaria o una serie de tiempo que sea estacionaria después de una o más diferenciaciones. La razón para suponer estacionariedad se explica de la siguiente manera:

> El objetivo de BJ [Box-Jenkins] es identificar y estimar un modelo estadístico que se interprete como generador de los datos muestrales. Entonces, si se va a pronosticar con este modelo estimado, debe suponerse que sus características son constantes a través del tiempo y, en particular, en periodos futuros. Así, la sencilla razón para requerir datos estacionarios es que todo modelo que se infiera a partir de estos datos pueda interpretarse como estacionario o estable en sí mismo, y proporcione, por consiguiente, una base válida para pronosticar.[6]

## 22.3 Metodología de Box-Jenkins (BJ)

La pregunta del millón de dólares es obvia: al ver una serie de tiempo, como la serie del PIB de Estados Unidos en la figura 21.1, ¿cómo sabemos si sigue un proceso AR puro (de ser así, cuál es el valor de $p$), un proceso MA puro (de ser así, cuál es el valor de $q$), un proceso ARMA (de ser así, cuáles son los valores de $p$ y $q$) o un proceso ARIMA, en cuyo caso se deben conocer los valores de $p$, $d$ y $q$? La metodología BJ resulta útil para responder la pregunta anterior. El método considera cuatro pasos:

**Paso 1. Identificación.** Es decir, encontrar los valores apropiados de $p$, $d$ y $q$. En seguida veremos la forma como el **correlograma** y el **correlograma parcial** ayudan en esta labor.

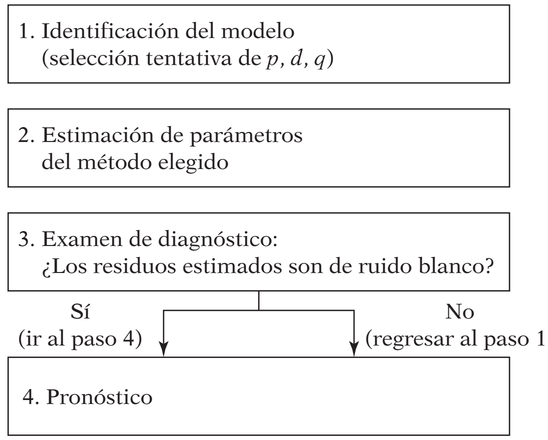
**Paso 2. Estimación.** Tras identificar los valores apropiados de $p$ y $q$, la siguiente etapa es estimar los parámetros de los términos autorregresivos y de promedios móviles incluidos en el modelo. Algunas veces, este cálculo se efectúa mediante mínimos cuadrados simples, pero otras hay que recurrir a métodos de estimación no lineal (en parámetros). Como esta labor se lleva a cabo ahora a través de rutinas en diversos paquetes estadísticos, en la práctica no es preciso preocuparse por los desarrollos matemáticos de la estimación; el estudiante interesado en el tema puede consultar las referencias.

**Paso 3. Examen de diagnóstico.** Después de seleccionar un modelo ARIMA particular y de estimar sus parámetros, tratamos de ver si el modelo seleccionado se ajusta a los datos en forma razonablemente buena, pues es posible que exista otro modelo ARIMA que también lo haga. Es por esto que el diseño de modelos ARIMA de Box-Jenkins es un arte más que una ciencia; se requiere gran habilidad para seleccionar el modelo ARIMA correcto. Una simple prueba del modelo seleccionado es ver si los residuales estimados a partir de este modelo son de ruido blanco; si lo son, aceptamos el ajuste particular; si no lo son, debemos empezar de nuevo. **Por tanto, la metodología BJ es un proceso iterativo** (figura 22.1).

---

[6] Michael Pokorny, *An Introduction to Econometrics*, Basil Blackwell, Nueva York, 1987, p. 343.

**FIGURA 22.1**
Metodología de Box-Jenkins.




**Paso 4. Pronóstico.** Una razón de la popularidad del proceso de construcción de modelos ARIMA es su éxito en el pronóstico. En muchos casos, los pronósticos obtenidos por este método son más confiables que los obtenidos de modelos econométricos tradicionales, en particular en el caso de pronósticos de corto plazo. Por supuesto, cada caso debe verificarse.

Luego de esta exposición general, examinaremos los cuatro pasos con cierto detalle. Durante el análisis utilizaremos los datos del PIB de la sección 21.1 para ilustrar los diversos puntos.

## 22.4 Identificación

Las herramientas principales en la identificación son la **función de autocorrelación (FAC)**, la **función de autocorrelación parcial (FACP)** y los **correlogramas** resultantes, que son simplemente los gráficos de FAC y de FACP respecto de la longitud del rezago.

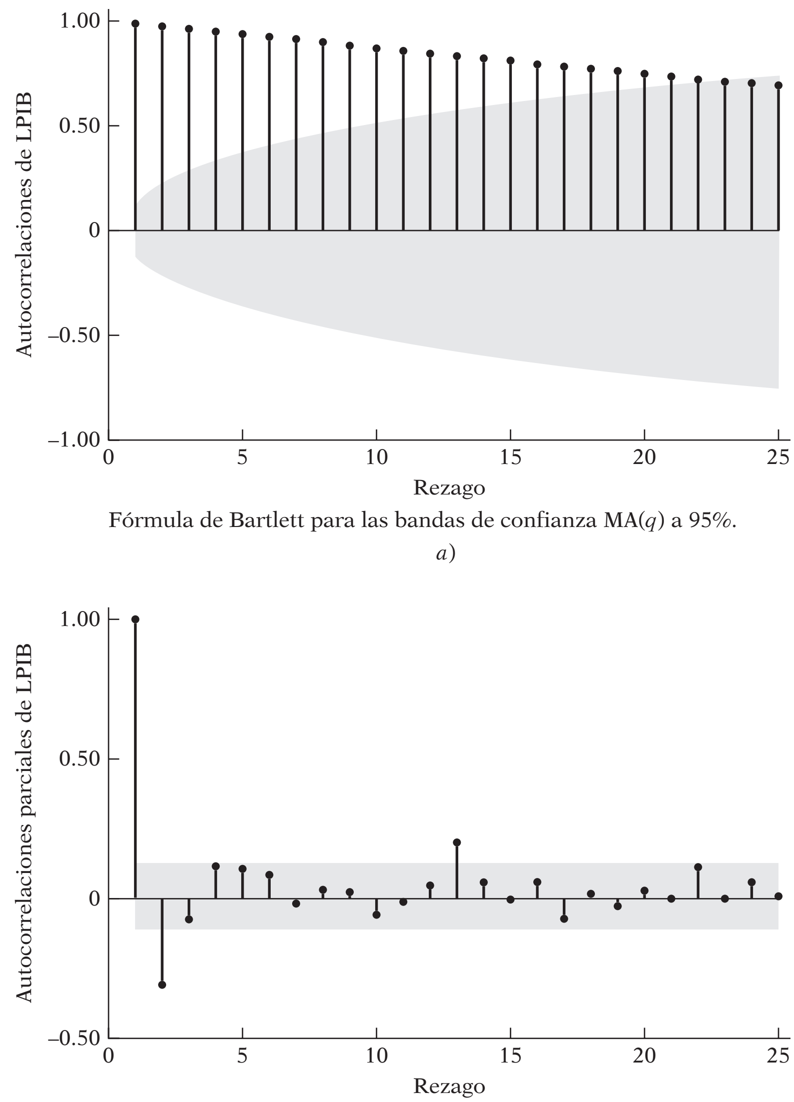
En el capítulo anterior definimos la FAC $\rho_k$ (poblacional) y la FAC $\hat{\rho}_k$ (muestral). El concepto de autocorrelación parcial es análogo al concepto de coeficiente de regresión parcial. En el modelo de regresión múltiple con $k$ variables, el $k$-ésimo coeficiente de regresión $\beta_k$ mide la tasa de cambio en el valor medio de la variable regresada ante un cambio unitario en la $k$-ésima regresora $X_k$, para mantener constante la influencia de todas las demás regresoras.

En forma similar, la **autocorrelación parcial** $\rho_{kk}$ mide la correlación entre observaciones (series de tiempo) separadas $k$ periodos y mantiene constantes las correlaciones en los rezagos intermedios (es decir, rezagos menores de $k$). En otras palabras, la autocorrelación parcial es la correlación entre $Y_t$ y $Y_{t-k}$ después de eliminar el efecto de las $Y$ intermedias.[7] En la sección 7.11 presentamos el concepto de correlación parcial en el contexto de regresión y vimos su relación con las correlaciones simples. Tales correlaciones parciales se calculan ahora mediante rutinas en la mayoría de los paquetes estadísticos.

En la figura 22.2 presentamos el correlograma y el correlograma parcial de la serie LPIB. En esta figura sobresalen dos hechos: primero, la FAC decrece muy lentamente; como se ve en la figura 21.8, la FAC hasta el rezago 23 es estadísticamente diferente de cero en un nivel individual, pues todas están fuera de los límites a 95% de confianza. Segundo, después del primer rezago, la FACP decrece en forma considerable y todas las FACP, después del primer rezago, son estadísticamente no significativas, salvo quizá por el rezago 13.

[7] En los datos de series de tiempo, una gran proporción de la correlación entre $Y_t$ y $Y_{t-k}$ puede deberse a sus correlaciones con los rezagos intermedios $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$. La correlación parcial $\rho_{kk}$ elimina la influencia de estas variables intermedias.

**FIGURA 22.2**
*a*) Correlograma y *b*) correlograma parcial de LPIB, Estados Unidos, I-1947 a IV-2007.



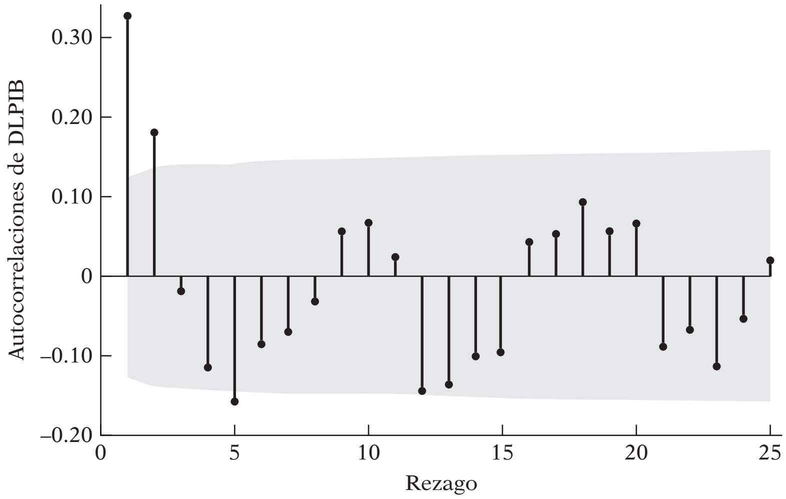

Fórmula de Bartlett para las bandas de confianza MA($q$) a 95%.
*a*)

Bandas de confianza a 95% [ee = 1/raíz cuadrada de ($n$)].
*b*)

Como la serie de tiempo LPIB de Estados Unidos es no estacionaria, debemos convertirla en estacionaria antes de aplicar la metodología de Box-Jenkins. En la figura 21.9 graficamos la serie de primeras diferencias del LPIB. En contraste con la figura 21.1, no se observa ninguna tendencia en esta serie, lo que sugiere que quizá la serie de tiempo del LPIB en primeras diferencias es estacionaria.[8] Una aplicación formal de la **prueba de raíz unitaria de Dickey-Fuller** muestra que así sucede, en efecto. Lo anterior también se visualiza mediante los correlogramas FAC y FACP estimados en los paneles *a*) y *b*) de la figura 22.3. Ahora tenemos un patrón muy diferente de FAC y de FACP. Las FACs en los rezagos 1, 2 y 5 parecen estadísticamente diferentes de cero; recuerde, del capítulo 21, que los límites de confianza aproximados a 95% para $\rho_k$ son $-0.2089$ y $+0.2089$. (*Nota:* Como analizamos en el capítulo 21, estos límites de confianza son asintóticos

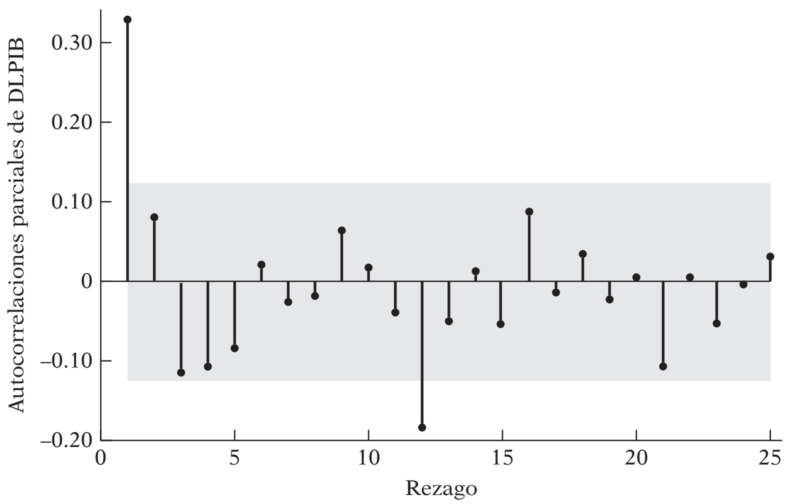
[8] Es difícil decir si la varianza de esta serie es estacionaria, en especial alrededor de 1979-1980. El embargo petrolero de 1979 y un cambio significativo en la política monetaria del Banco de la Reserva Federal en el mismo año pueden tener algo que ver con esta dificultad.

**FIGURA 22.3**
*a*) Correlograma y *b*) correlograma parcial de las primeras diferencias de LPIB, Estados Unidos, I-1947 a IV-2007.




Fórmula de Bartlett para las bandas de confianza MA($q$) a 95%.

*a*)




Bandas de confianza a 95% [ee = 1/raíz cuadrada de ($n$)].

*b*)

y, por tanto, se consideran aproximados.) Pero, en todos los demás rezagos, no son estadísticamente diferentes de cero. Para las autocorrelaciones parciales, sólo los rezagos 1 y 12 parecen estadísticamente diferentes de cero.

Ahora, ¿cómo permiten los correlogramas de la figura 22.3 encontrar el patrón ARMA de la serie de tiempo del PIB? (*Nota:* Sólo consideraremos la serie del PIB en primeras diferencias por ser estacionaria.) Una forma de lograrlo es considerar la FAC y la FACP, y los correlogramas asociados de un número seleccionado de procesos ARMA, como AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARIMA(2,2), y así sucesivamente. Como cada proceso estocástico presenta patrones habituales de FAC y de FACP, si la serie de tiempo en estudio se ajusta a alguno de estos patrones, la podemos identificar con tal proceso. Desde luego, será necesario aplicar pruebas de diagnóstico para determinar si el modelo seleccionado ARMA es razonablemente preciso.

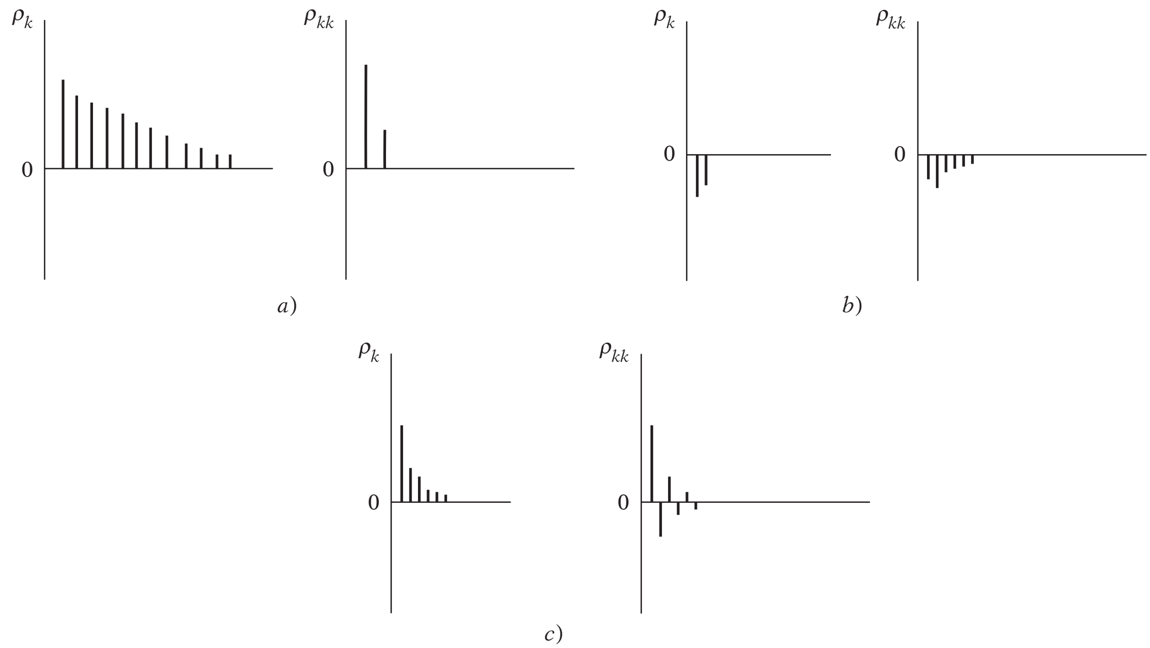
El estudio de las propiedades de los diversos procesos estándar ARIMA consumiría mucho espacio. En su lugar, ofrecemos lineamientos generales (tabla 22.1); en las referencias hay detalles de los diversos procesos estocásticos.

**TABLA 22.1**
**Patrones teóricos de FAC y FACP**

| Tipo de modelo | Patrón típico de FAC | Patrón típico de FACP |
|---|---|---|
| AR($p$) | Disminuye exponencialmente o con un patrón de onda sinusoide achatada o ambos. | Picos significativos en los rezagos $q$ |
| MA($q$) | Picos significativos en los rezagos $q$ | Decrece exponencialmente |
| ARMA($p$, $q$) | Decrece exponencialmente | Decrece exponencialmente |

*Nota:* Los términos *decrecimiento exponencial* y *decrecimiento geométrico* significan lo mismo (recuerde el análisis sobre el rezago distribuido de Koyck).

**FIGURA 22.4** FAC y FACP de procesos estocásticos seleccionados: *a*) AR(2): $\alpha_1 = 0.5$, $\alpha_2 = 0.3$; *b*) MA(2): $\beta_1 = 0.5$, $\beta_2 = 0.3$; *c*) ARMA(1,1): $\alpha_1 = 0.5$, $\beta_1 = 0.5$.




Observe que las FAC y las FACP de los procesos AR($p$) y MA($q$) tienen patrones opuestos; en el caso AR($p$), la FAC decrece geométrica o exponencialmente, pero la FACP se corta después de cierto número de rezagos, mientras que sucede lo opuesto a un proceso MA($q$).

Geométricamente, estos patrones se muestran en la figura 22.4.

### *Advertencia*

Como en la práctica no se observan las FAC y FACP teóricas y se depende, por tanto, de sus aproximaciones muestrales, las FAC y FACP estimadas no concordarán exactamente con sus contrapartes teóricas. Buscamos una similitud entre las FAC y las FACP teóricas y muestrales de manera que señalen la dirección correcta en la construcción de los modelos ARIMA. Es por esto que la elaboración de modelos ARIMA requiere gran habilidad, lo cual, desde luego, se obtiene con la práctica.

### *Identificación ARIMA del PIB de Estados Unidos*

Al considerar de nuevo el correlograma y el correlograma parcial de LPIB estacionario (después de la primera diferenciación) de Estados Unidos de I-1947 a IV-2007 de la figura 22.3, ¿qué observamos?

Recordemos que la FAC y la FACP son cantidades muestrales, y no tenemos un patrón decente como los propuestos en la tabla 22.1. Las autocorrelaciones (panel *a*) decrecen en los primeros dos rezagos y luego, con excepción del rezago 5, los restantes no son estadísticamente diferentes

de cero (el área gris en las figuras da los límites de confianza aproximados a 95%). Las autocorrelaciones parciales (panel *b*) con picos en los rezagos 1 y 12 parecen estadísticamente significativas, pero el resto no; si el coeficiente de correlación parcial fuera significativo solamente en el rezago 1, podríamos haberlo identificado como un modelo AR(1). Supongamos por tanto que el proceso que generó la serie LPIB (en primeras diferencias) es un proceso MA(2). Tenga en cuenta que, a menos que la FAC y la FACP no estén bien definidas, será difícil elegir un modelo sin ensayo y error. Se invita al lector a probar otros modelos ARIMA con la serie LPIB en primeras diferencias.

## 22.5 Estimación del modelo ARIMA

Sea $Y_t^*$ las primeras diferencias del logaritmo del PIB de Estados Unidos. Entonces, el modelo MA identificado tentativamente es

$$Y_t^* = \mu + \beta_1 u_{t-1} + \beta_2 u_{t-2} \tag{22.5.1}$$

Con MINITAB obtuvimos las siguientes estimaciones:

$$\begin{aligned} \hat{Y}_t^* &= \ 0.00822 + 0.2918u_{t-1} + 0.2024u_{t-2} \\ \text{ee} &= (0.00088) \quad (0.0633) \qquad (0.0634) \\ t &= (9.32) \qquad (4.61) \qquad (3.20) \\ & \qquad R^2 = 0.1217 \qquad d = 1.9705 \end{aligned} \tag{22.5.2}$$

Dejamos como ejercicio al lector estimar otros modelos ARIMA para la serie LPIB en primeras diferencias.

## 22.6 Verificación de diagnóstico

¿Cómo sabemos que el modelo en (22.5.2) es un ajuste razonable a los datos? Un diagnóstico simple es obtener los residuos de (22.5.2) y obtener la FAC y la FACP de estos residuos, por ejemplo, hasta el rezago 25. Las FAC y FACP estimadas se muestran en la figura 22.5. Como vemos en esta figura, ninguna de las autocorrelaciones y de las autocorrelaciones parciales es estadísticamente significativa de manera individual. Tampoco lo es la suma de las 25 autocorrelaciones elevadas al cuadrado, como indican los estadísticos Q de Box-Pierce y LB de Ljung-Box (véase el capítulo 21). En otras palabras, los correlogramas de la autocorrelación y de la autocorrelación parcial dan la impresión de que los residuos estimados de (22.5.2) son puramente aleatorios. Por tanto, quizá no sea necesario buscar otro modelo ARIMA.

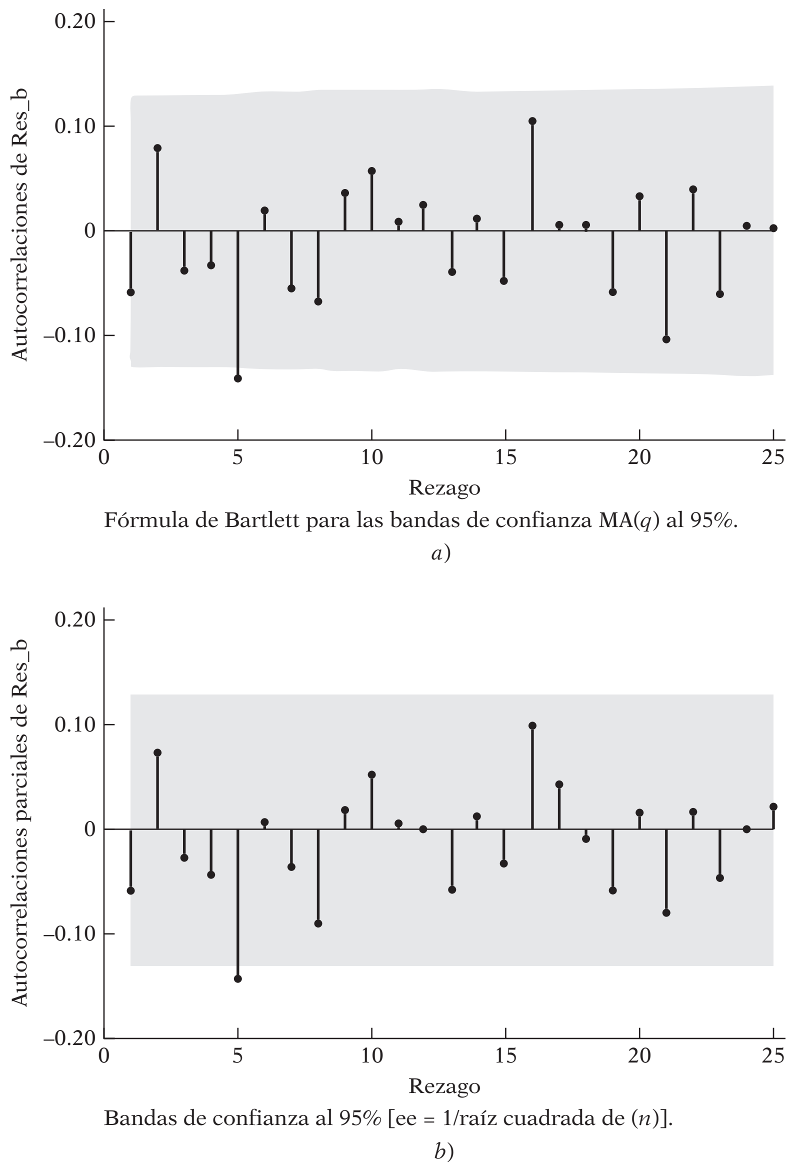
## 22.7 Pronóstico

Recuerde que los datos del PIB son para el periodo I-1947 a IV-2007. Suponga, con base en el modelo (22.5.2), que deseamos pronosticar el LPIB para el primero de los cuatro trimestres de 2008. Pero, en (22.5.2), la variable dependiente es el *cambio* en el LPIB en el trimestre anterior. Por consiguiente, si utilizamos (22.5.2), lo que obtenemos son los pronósticos de los cambios en el LPIB entre el primer trimestre de 2008 y el cuarto trimestre de 2007, el segundo trimestre de 2008 y el primer trimestre de 2008, etcétera.

Para obtener el pronóstico de nivel del LPIB en lugar de sus cambios, podemos "deshacer" la transformación de primeras diferencias con que obtuvimos los cambios. (Más técnicamente, *integramos* la serie de primeras diferencias.) Así, para obtener el valor de pronóstico del LPIB (no de ΔPIB) para I-2008, reescribimos el modelo (22.5.1) como

$$Y_{2008\text{-I}} - Y_{2007\text{-IV}} = \mu + \beta_1 u_{2007\text{-IV}} + \beta_2 u_{2007\text{-III}} + u_{2008\text{-I}} \tag{22.7.1}$$

**FIGURA 22.5**
*a*) Correlograma y *b*) correlograma parcial de los residuos del modelo MA(2) para las primeras diferencias de LPIB, Estados Unidos, 1947-I a 2007-IV.




Es decir,

$$Y_{2008\text{-I}} = \mu + \beta_1 u_{2007\text{-IV}} + \beta_2 u_{2007\text{-III}} + u_{2008\text{-I}} + Y_{2007\text{-IV}} \tag{22.7.2}$$

Los valores de $\mu$, $\beta_1$ y $\beta_2$ ya se conocen de la regresión estimada (22.5.2). Se supone que el valor de $u_{2008\text{-I}}$ es cero (¿por qué?). Por consiguiente, obtenemos fácilmente el valor de pronóstico de $Y_{2008\text{-I}}$. La estimación numérica de este valor de pronóstico es:[9]

$$\begin{aligned} Y_{2008\text{-I}} &= 0.00822 + (0.2918)u_{2007\text{-IV}} + (0.2024)(u_{2007\text{-III}}) + Y_{2007\text{-IV}} \\ &= 0.00822 + (0.2918)(0.00853) + (0.2024)(-0.00399) + 9.3653 \\ &= 9.3741 (\text{aprox.}) \end{aligned}$$

[9] Aunque los paquetes de software comunes y corrientes hacen este cálculo habitualmente, presentamos los cálculos detallados para ilustrar el procedimiento.

Así, el valor de pronóstico de LPIB para I-2008 es de alrededor de 9.3741, que equivale a unos 11 779 millones de dólares (de 2000). A propósito, el valor observado del PIB real para I-2008 fue de 11 693.09 millones de dólares; el error de pronóstico fue una sobreestimación de 86 000 millones de dólares.

## 22.8 Otros aspectos de la metodología BJ

En los párrafos anteriores presentamos sólo una introducción general al diseño de modelos BJ. Hay muchos aspectos de esta metodología no considerados por falta de espacio, como la **estacionalidad**. Muchas series de tiempo presentan un comportamiento estacional, como las ventas de tiendas departamentales en días festivos, el consumo estacional de helado, los viajes durante días festivos nacionales, etc. Si, por ejemplo, disponemos de la información trimestral de ventas de las tiendas departamentales, estas cifras mostrarán picos en el cuarto trimestre. En tales situaciones, es posible eliminar la influencia estacional al tomar diferencias trimestrales de orden cuatro de las cifras de ventas y luego decidir qué clase de modelo ARIMA ajustar.

Analizamos una serie de tiempo a la vez; sin embargo, nada impide extender la metodología BJ al estudio simultáneo de dos o más series de tiempo. Una revisión de tal tema se saldría del alcance de este libro. El lector interesado puede consultar las referencias.[10] No obstante, en la siguiente sección analizaremos este tema en el contexto de lo que se conoce como vectores autorregresivos.

## 22.9 Vectores autorregresivos (VAR)

De los capítulos 18 a 20 consideramos los modelos de ecuaciones simultáneas o estructurales. En esos modelos, algunas variables se tratan como endógenas y otras como exógenas o predeterminadas (exógenas y endógenas rezagadas). Antes de estimar tales modelos debemos estar seguros de que las ecuaciones en el sistema estén identificadas (en forma exacta o sobreidentificadas). Esta identificación a menudo se logra si suponemos que algunas variables predeterminadas sólo están presentes en ciertas ecuaciones. Esta decisión a menudo es subjetiva y ha recibido fuertes críticas de Christopher Sims.[11]

De acuerdo con Sims, si hay verdadera simultaneidad entre un conjunto de variables, todas deben tratarse en igualdad de condiciones: no debe haber ninguna distinción *a priori* entre variables endógenas y exógenas. Es en este contexto que Sims desarrolló su modelo **VAR**.

Las semillas de este modelo se habían sembrado ya en la prueba de causalidad de Granger, estudiada en el capítulo 17. En las ecuaciones (17.14.1) y (17.14.2), que explican el LPIB actual en términos de la oferta monetaria rezagada y del PIB rezagado, y la oferta monetaria actual en términos de la oferta monetaria rezagada y del LPIB rezagado, tratamos esencialmente al PIB y a la oferta monetaria como un par de variables endógenas. No hay variables exógenas en este sistema.

De igual forma, en el ejemplo 17.13 analizamos la naturaleza de la causalidad entre dinero y tasa de interés en Canadá. En la ecuación del dinero, sólo aparecen los valores rezagados del dinero y de la tasa de interés; asimismo, en la ecuación de la tasa de interés, únicamente se presentan los valores rezagados de la tasa de interés y del dinero.

Estos dos ejemplos son ilustraciones de **modelos de vectores autorregresivos**; el término "autorregresivo" se refiere a la aparición del valor rezagado de la variable dependiente en el lado derecho, y el término "vector" se atribuye a que tratamos con un vector de dos (o más) variables.

[10] Para un tratamiento fácil de entender sobre este tema, véase Terence C. Mills, *op. cit.*, parte III.

[11] C.A. Sims, "Macroeconomics and Reality", en *Econometrica*, vol. 48, 1980, pp. 1-48.

## Estimación de VAR

De vuelta con el ejemplo del dinero y la tasa de interés para Canadá, vimos que, cuando introdujimos seis rezagos de cada variable como regresoras, no pudimos rechazar la hipótesis de que existía una causalidad bilateral entre el dinero ($M_1$) y la tasa de interés $R$ (tasa de interés corporativa a 90 días). Es decir, $M_1$ afecta a $R$ y ésta a su vez influye en $M_1$. Este tipo de situaciones son ideales para aplicar el método VAR.

A fin de explicar cómo se estima VAR, continuaremos con el ejemplo anterior. Por el momento, suponga que cada ecuación contiene $k$ valores rezagados de $M$ (como se mide $M_1$) y $R$. En este caso, estimamos cada una de las siguientes ecuaciones mediante MCO.[12]

$$M_{1t} = \alpha + \sum_{j=1}^{k} \beta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{1t} \tag{22.9.1}$$

$$R_t = \alpha' + \sum_{j=1}^{k} \theta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{2t} \tag{22.9.2}$$

donde las $u$ son los términos de error estocástico, llamados **impulsos**, **innovaciones** o **choques** en el lenguaje de VAR.

Antes de calcular (22.9.1) y (22.9.2) tenemos que decidir la longitud máxima del rezago $k$. Esto significa una cuestión empírica. Se tienen 40 observaciones en total. La inclusión de muchos términos rezagados consumirá muchos grados de libertad, para no mencionar la posible aparición de la multicolinealidad. Agregar muy pocos rezagos provoca errores de especificación. Una forma de decidir esta cuestión es utilizar criterios como el de Akaike o el de Schwarz, para después elegir el modelo que proporcione los valores más bajos de dichos criterios. En definitiva, no podemos evitar cierto grado de ensayo y error.

Con el objeto de ilustrar el procedimiento, al principio utilizamos cuatro rezagos ($k = 4$) de cada variable y mediante el programa *EViews 6* obtuvimos las estimaciones de los parámetros de las dos ecuaciones anteriores, los cuales se proporcionan en la tabla 22.2. Observe que, a pesar de que la muestra abarca desde el primer trimestre de 1979 hasta el cuarto de 1988, utilizamos la muestra para el periodo comprendido entre el primer trimestre de 1980 y el cuarto de 1987, y conservamos las últimas cuatro observaciones para verificar la precisión del pronóstico del VAR ajustado.

En vista de que las ecuaciones anteriores son regresiones de MCO, el resultado de la regresión de la tabla 22.2 debe interpretarse de la manera usual. Por supuesto, con diversos rezagos de las mismas variables, cada coeficiente estimado no será estadísticamente significativo, quizá debido a la multicolinealidad. Pero en forma colectiva, quizá sean significativos respecto de la prueba $F$ estándar.

Examinemos los resultados mostrados en la tabla 22.2. Primero consideramos la regresión $M_1$. En forma individual, sólo $M_1$ en el rezago 1, y $R$, en los rezagos 1 y 2, son estadísticamente significativos. Pero el valor $F$ es tan alto que no podemos rechazar la hipótesis de que de manera colectiva todos los términos de rezago son estadísticamente significativos. Al concentrar la atención sobre la regresión de la tasa de interés, observamos que los cuatro rezagos de los términos monetarios son estadísticamente significativos considerados de modo individual (en un nivel de 10% o superior), en tanto que sólo la variable de la tasa de interés con rezago de un periodo resulta significativa.

Para propósitos comparativos, en la tabla 22.3 se presentan los resultados VAR basados en sólo dos rezagos de cada variable endógena. Aquí apreciará que, en la regresión sobre el dinero, la variable dinero con rezago de un periodo y los términos de la tasa de interés rezagada son

[12] Se puede utilizar la técnica **SURE** (*seemingly unrelated regression*) para estimar las dos ecuaciones de manera simultánea. No obstante, como cada regresión contiene el mismo número de variables endógenas rezagadas, la estimación por MCO de cada ecuación por separado suministra estimaciones idénticas (y eficientes).

**TABLA 22.2**
**Estimaciones de vectores autorregresivos con base en cuatro rezagos**

Muestra (ajustada): primer trimestre de 1980 a cuarto trimestre de 1987
Observaciones incluidas: 32 después de ajustar los extremos
Los errores estándar están entre paréntesis y el estadístico *t* entre corchetes

| | $M_1$ | $R$ |
|---|---|---|
| $M_1$ (−1) | 1.076737 (0.20174) [5.33733] | 0.001282 (0.00067) [1.90083] |
| $M_1$ (−2) | 0.173433 (0.31444) [0.55157] | −0.002140 (0.00105) [−2.03584] |
| $M_1$ (−3) | −0.366465 (0.34687) [−1.05648] | 0.002176 (0.00116) [1.87699] |
| $M_1$ (−4) | 0.077602 (0.20789) [0.37329] | −0.001479 (0.00069) [−2.12855] |
| $R$ (−1) | −275.0293 (57.2174) [−4.80675] | 1.139310 (0.19127) [5.95670] |
| $R$ (−2) | 227.1750 (95.3947) [2.38142] | −0.309053 (0.31888) [−0.96917] |
| $R$ (−3) | 8.511851 (96.9176) [0.08783] | 0.052361 (0.32397) [0.16162] |
| $R$ (−4) | −50.19926 (64.7554) [−0.77521] | 0.001076 (0.21646) [0.00497] |
| $C$ | 2 413.827 (1 622.65) [1.48759] | 4.919000 (5.42416) [0.90687] |
| $R^2$ | 0.988154 | 0.852890 |
| $R^2$ ajustada | 0.984034 | 0.801721 |
| Suma de cuadrados residual | 4 820 241. | 53.86233 |
| Ecuación del ee | 457.7944 | 1.530307 |
| Estadístico $F$ | 239.8315 | 16.66815 |
| Log verosimilitud | −236.1676 | −53.73716 |
| Akaike A/C | 15.32298 | 3.921073 |
| Schwarz SC | 15.73521 | 4.333311 |
| Media de la variable dependiente | 28 514.53 | 11.67292 |
| DE de la variable dependiente | 3 623.058 | 3.436688 |
| Determinante de covarianza residual | 490 782.3 | |
| Log verosimilitud (gl ajustados) | −300.4722 | |
| Criterio de información de Akaike | 19.90451 | |
| Criterio de Schwarz | 20.72899 | |

estadísticamente significativos en lo individual. En la regresión sobre la tasa de interés, los términos dinero rezagado (en casi un nivel de 5%) e interés rezagado resultan individualmente significativos.

Si tenemos que hacer una elección entre el modelo de la tabla 22.2 y el de la 22.3, ¿cuál escogería? Los valores de información Akaike y Schwarz para el modelo de la tabla 22.2 son 15.32 y 15.73, respectivamente, en tanto que los correspondientes a la tabla 22.3 son 15.10 y 15.33. Como mientras más bajos sean los valores de los estadísticos Akaike y Schwarz mejor será el modelo, parece preferible el modelo más parco de la tabla 22.3. También consideramos seis rezagos de cada variable endógena y descubrimos que los valores de los estadísticos Akaike y Schwarz fueron de 15.37 y 15.98, respectivamente. De nuevo, la elección parece el modelo con dos términos rezagados de cada variable endógena, es decir, el modelo de la tabla 22.3.

## Pronóstico con el modelo VAR

Suponga que elegimos el modelo de la tabla 22.3 para predecir los valores de $M_1$ y $R$. Recuerde que los datos abarcan el periodo del primer trimestre de 1979 al cuarto de 1988, pero no utilizamos los valores para 1988 al estimar los modelos VAR. Ahora suponga que deseamos pronosticar el valor de $M_1$ para I-1988, es decir, el primer trimestre de 1988. Obtenemos el valor pronosticado para I-1988 de la siguiente forma:

$$\begin{aligned}\hat{M}_{1988-\text{I}} = {} & 1\,451.977 + 1.0375 M_{1987-\text{IV}} - 0.0446 M_{1987-\text{III}} \\ & - 234.8850 R_{1987-\text{IV}} + 160.1560 R_{1987-\text{III}}\end{aligned}$$

**TABLA 22.3**
**Estimaciones de vectores autorregresivos con base en dos rezagos**

Muestra (ajustada): tercer trimestre de 1979 al cuarto trimestre de 1987
Observaciones incluidas: 34 después de ajustar los extremos
Los errores estándar están entre paréntesis y el estadístico *t* entre corchetes

| | $M_1$ | $R$ |
|---|---|---|
| $M_1$ (−1) | 1.037537 (0.16048) [6.46509] | 0.001091 (0.00059) [1.85825] |
| $M_1$ (−2) | −0.044661 (0.15591) [−0.28646] | −0.001255 (0.00057) [−2.19871] |
| $R$ (−1) | −234.8850 (45.5224) [−5.15977] | 1.069081 (0.16660) [6.41708] |
| $R$ (−2) | 160.1560 (48.5283) [3.30026] | −0.223364 (0.17760) [−1.25768] |
| $C$ | 1 451.977 (1 185.59) [1.22468] | 5.796434 (4.33894) [1.33591] |
| $R^2$ | 0.988198 | 0.806660 |
| $R^2$ ajustada | 0.986571 | 0.779993 |
| Suma de cuadrados residual | 5 373 510. | 71.97054 |
| Ecuación del ee | 430.4573 | 1.575355 |
| Estadístico $F$ | 607.0720 | 30.24878 |
| Log verosimilitud | −251.7446 | −60.99215 |
| Akaike A/C | 15.10263 | 3.881891 |
| Schwarz SC | 15.32709 | 4.106356 |
| Media de la variable dependiente | 28 216.26 | 11.75049 |
| DE de la variable dependiente | 3 714.506 | 3.358613 |
| Determinante de covarianza residual | 458 485.4 | |
| Log verosimilitud (gl ajustados) | −318.0944 | |
| Criterio de información de Akaike | 19.29967 | |
| Criterio de Schwarz | 19.74860 | |

donde los valores de los coeficientes se obtuvieron de la tabla 22.3. Ahora, con los valores apropiados de $M_1$ y $R$ de la tabla 17.5, podemos afirmar que el valor pronosticado del dinero para el primer trimestre de 1988 es de 36 996 (millones de dólares canadienses). El valor real de $M_1$ para el primer trimestre de 1988 fue de 36 480, lo que significa que el modelo pronosticó en exceso el valor real por casi 516 (millones de dólares), cantidad que representa casi 1.4% del valor real de $M_1$ para I-1988. Desde luego, estas estimaciones cambian en función de la cantidad de valores rezagados que consideremos en el modelo VAR. Dejamos como ejercicio para el lector pronosticar el valor de $R$ para el primer trimestre de 1988 y compararlo con su valor real para dicho periodo.

## VAR y causalidad

Recordará que ya estudiamos el tema de la causalidad en el capítulo 17. Ahí estudiamos las pruebas de Granger y Sims de la causalidad. ¿Existe alguna conexión entre el modelo VAR y la causalidad? En el capítulo 17 (sección 17.14) vimos que, hasta el nivel de 2, 4 y 6 rezagos, había una causalidad bilateral entre $M_1$ y $R$, pero que en el rezago 8 no existía causalidad entre las dos variables. Por consiguiente, los resultados son ambiguos. Ahora recordará, del capítulo 21, el teorema de representación de Granger. Una implicación de este teorema es que si dos variables, por ejemplo $X_t$ y $Y_t$, están cointegradas y cada una es individualmente $I(1)$, es decir, es integrada de orden 1 (por ejemplo, cada una es no estacionaria individualmente), $X_t$ debe causar (según el criterio de Granger) a $Y_t$, o $Y_t$ debe causar (según el criterio de Granger) a $X_t$.

En el ejemplo ilustrativo, esto significa que $M_1$ y $R$ son individualmente $I(1)$ pero están cointegradas, entonces $M_1$ debe causar (según el criterio de Granger) a $R$ o $R$ debe causar (según el criterio de Granger) a $M_1$. Esto significa que debemos averiguar primero si ambas variables son individualmente $I(1)$ y luego si están cointegradas. De no ser así, todo el asunto de la causalidad

se convertiría en una ficción. En el ejercicio 22.22 se pide al lector determinar si las dos variables son no estacionarias pero están cointegradas. Si efectúa el ejercicio, descubrirá una *débil* evidencia de cointegración entre $M_1$ y $R$, razón por la cual las pruebas de causalidad analizadas en la sección 17.14 eran erróneas.

## Algunos problemas en la creación de modelos VAR

Los defensores de VAR destacan las siguientes virtudes del método: 1) El método es simple; no es preciso preocuparse por determinar cuáles variables son endógenas y cuáles son exógenas. Todas las variables en VAR son endógenas.[13] 2) La estimación es simple, es decir, el método usual de MCO es aplicable a cada ecuación por separado. 3) Las predicciones obtenidas mediante este método son en muchos casos mejores que las obtenidas con modelos de ecuaciones simultáneas más complejos.[14]

No obstante, los críticos de los modelos VAR señalan los siguientes problemas:

1. A diferencia de los modelos de ecuaciones simultáneas, un modelo VAR es *ateórico* porque utiliza menos información previa. Recuerde que en los modelos de ecuaciones simultáneas, la exclusión o inclusión de ciertas variables desempeña un papel crucial en la identificación del modelo.

2. Debido a su acento en el pronóstico, los modelos VAR son menos apropiados para el análisis de políticas.

3. El mayor desafío práctico en el diseño de modelos VAR es seleccionar la longitud apropiada del rezago. Suponga que tenemos un modelo VAR de tres variables y decidimos incluir ocho rezagos de cada variable en cada ecuación. Tendremos 24 parámetros rezagados en cada ecuación más el término constante, para un total de 25 parámetros. A menos que el tamaño de la muestra sea grande, la estimación de tantos parámetros consumirá muchos grados de libertad, con todos los problemas asociados a esto.[15]

4. En sentido estricto, en un modelo VAR de $m$ variables, todas las $m$ variables deben ser estacionarias (en forma conjunta). Si no es así, tendremos que transformar la información en forma apropiada (por ejemplo, mediante una primera diferenciación). Como menciona Harvey, los resultados de la información transformada pueden ser no satisfactorios. Afirma además: "El método usual adoptado por los *aficionados* al VAR es, por consiguiente, trabajar en niveles, aunque algunas de estas series no sean estacionarias. En este caso es importante reconocer el efecto de las raíces unitarias en la distribución de los estimadores".[16] Peor aún, si el modelo contiene una mezcla de variables $I(0)$ e $I(1)$, es decir, una mezcla de variables estacionarias y no estacionarias, no será fácil transformar la información.

Sin embargo, Cuthbertson sostiene: "El análisis de cointegración indica que un VAR *sólo* en primeras diferencias está mal especificado si hay algunos vectores cointegrantes presentes entre la serie $I(1)$. Dicho de otro modo, un VAR exclusivamente en primeras diferencias omite

[13] Algunas veces se incluyen variables puramente exógenas para dar cabida a factores estacionales y de tendencia.

[14] Véase, por ejemplo, T. Kinal y J.B. Ratner, "Regional Forecasting Models with Vector Autoregression: The Case of New York State", documento de trabajo # 155, Departamento de Economía, State University of New York at Albany, 1982.

[15] Si tenemos un modelo VAR de $m$ ecuaciones con $p$ valores rezagados de las $m$ variables, en total tenemos que estimar $(m + pm^2)$ parámetros.

[16] Andrew Harvey, *The Econometric Analysis of Time Series*, The MIT Press, 2a. ed., Cambridge, Massachusetts, 1990, p. 83.

variables estacionarias potencialmente importantes (es decir, la corrección de errores, vectores cointegrantes) y, por consiguiente, las estimaciones de los parámetros pueden padecer el sesgo de variables omitidas".[17]

5. Como con frecuencia es difícil interpretar los coeficientes individuales estimados en los modelos VAR, los practicantes de esta técnica a menudo estiman la llamada **función de impulso-respuesta (FIR)**. La FIR estudia la respuesta de la variable dependiente en el sistema VAR ante choques en los términos de error, como $u_1$ y $u_2$ en las ecuaciones (22.9.1) y (22.9.2). Suponga que $u_1$ en la ecuación $M_1$ aumenta en el valor de una desviación estándar. Ese choque o cambio modificará a $M_1$ tanto en el periodo actual como en periodos futuros. Pero, como $M_1$ aparece en la regresión de $R$, el cambio en $u_1$ afectará también a $R$. En forma similar, un cambio de una desviación estándar en $u_2$ de la ecuación de $R$ tendrá un efecto sobre $M_1$. La FIR estudia el impacto de tales choques durante varios periodos en el futuro. Aunque los investigadores cuestionan la utilidad del análisis FIR, es la pieza central en el análisis VAR.[18]

Para comparar el desempeño de VAR con otras técnicas de pronóstico, el lector puede consultar las referencias.[19]

## Una aplicación de VAR: un modelo VAR de la economía de Texas

Para probar el refrán convencional: "Por donde va la mancha de petróleo se extiende la economía de Texas", Thomas Fomby y Joseph Hirschberg elaboraron un modelo VAR de tres variables de la economía de Texas de I-1974 a I-1988.[20] Las tres variables fueron: 1) cambio porcentual del precio real del petróleo, 2) cambio porcentual del empleo no agrícola de Texas y 3) cambio porcentual del empleo no agrícola en el resto de Estados Unidos. Los autores introdujeron el término constante y dos valores rezagados de cada variable en cada ecuación. Por consiguiente, el número de parámetros estimados en cada ecuación fue de siete. En la tabla 22.4 se presentan los resultados de la estimación por MCO del modelo VAR. Las pruebas $F$ de esta tabla deben corroborar la hipótesis de que, en forma conjunta, los diversos coeficientes rezagados son cero. Por tanto, la prueba $F$ para la variable $x$ (cambio porcentual del precio real del petróleo) muestra que los dos términos rezagados de $x$ son estadísticamente diferentes de cero; la probabilidad de obtener un valor $F$ de 12.5536, según la hipótesis nula de que ambos son simultáneamente iguales a cero, es muy baja, alrededor de 0.00004. Por otra parte, en forma conjunta, los dos valores $y$ rezagados (cambio porcentual en el empleo no agrícola de Texas) no son significativamente diferentes de cero para explicar $x$; el valor $F$ es sólo 1.36. Todos los demás estadísticos $F$ deben interpretarse en forma similar.

Con base en éstos y otros resultados presentados en su trabajo, Fomby y Hirschberg concluyen que el refrán sobre la economía de Texas no es del todo preciso, pues, después de la inestabilidad inicial resultante de las crisis del petróleo de los países de la OPEP, la economía de Texas es ahora menos dependiente de las fluctuaciones en el precio del petróleo.

[17] Keith Cuthbertson, *Quantitative Financial Economics: Stocks, Bonds and Foreigh Exchange,* John Wiley & Sons, Nueva York, 2002, p. 436.

[18] D.E. Runkle, "Vector Autoregression and Reality", *Journal of Business and Economic Statistics*, vol. 5, 1987, pp. 437-454.

[19] S. McNees, "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts", *Journal of Business and Economic Statistics*, vol. 4, 1986, pp. 5-15; E. Mahmoud, "Accuracy in Forecasting: A Survey", *Journal of Forecasting*, vol. 3, 1984, pp. 139-159.

[20] Thomas B. Fomby y Joseph G. Hiarschberg, "Texas in Transition: Dependence on Oil and the National Economy", en *Economic Review*, Federal Reserve Bank of Dallas, enero de 1989, pp. 11-28.

**TABLA 22.4**
**Resultados de la estimación del sistema VAR de segundo orden* de Texas: I-1974 a I-1988**

Fuente: *Economic Review*, Federal Reserve Bank of Dallas, enero de 1989, p. 21.

Variable dependiente: *x* (cambio porcentual del precio real del petróleo)

| Variable | Rezago | Coeficiente | Error estándar | Nivel de significancia |
|---|---|---|---|---|
| *x* | 1 | 0.7054 | 0.1409 | 0.8305E-5 |
| *x* | 2 | −0.3351 | 0.1500 | 0.3027E-1 |
| *y* | 1 | −1.3525 | 2.7013 | 0.6189 |
| *y* | 2 | 3.4371 | 2.4344 | 0.1645 |
| *z* | 1 | 3.4566 | 2.8048 | 0.2239 |
| *z* | 2 | −4.8703 | 2.7500 | 0.8304E-1 |
| Constante | 0 | −0.9983E-2 | 0.1696E-1 | 0.5589 |

$\bar{R}^2 = 0.2982$; $Q(21) = 8.2618$ $(P = 0.9939)$

Pruebas de significancia conjunta, variable dependiente = *x*

| Variable | Estadístico *F* | Nivel de significancia |
|---|---|---|
| *x* | 12.5536 | 0.4283E-4 |
| *y* | 1.3646 | 0.2654 |
| *z* | 1.5693 | 0.2188 |

Variable dependiente: *y* (cambio porcentual del empleo no agrícola de Texas)

| Variable | Rezago | Coeficiente | Error estándar | Nivel de significancia |
|---|---|---|---|---|
| *x* | 1 | 0.2228E-1 | 0.8759E-2 | 0.1430E-1 |
| *x* | 2 | −0.1883E-2 | 0.9322E-2 | 0.8407 |
| *y* | 1 | 0.6462 | 0.1678 | 0.3554E-3 |
| *y* | 2 | 0.4234E-1 | 0.1512 | 0.7807 |
| *z* | 1 | 0.2655 | 0.1742 | 0.1342 |
| *z* | 2 | −0.1715 | 0.1708 | 0.3205 |
| Constante | 0 | −0.1602E-2 | 0.1053E-1 | 0.1351 |

$\bar{R}^2 = 0.6316$; $Q(21) = 21.5900$ $(P = 0.4234)$

Pruebas de significancia conjunta, variable dependiente = *y*

| Variable | Estadístico *F* | Nivel de significancia |
|---|---|---|
| *x* | 3.6283 | 0.3424E-4 |
| *y* | 19.1440 | 0.8287E-6 |
| *z* | 1.1684 | 0.3197 |

Variable dependiente: *z* (cambio porcentual del empleo no agrícola en el resto de Estados Unidos)

| Variable | Rezago | Coeficiente | Error estándar | Nivel de significancia |
|---|---|---|---|---|
| *x* | 1 | −0.8330E-2 | 0.6849E-2 | 0.2299 |
| *x* | 2 | 0.3635E-2 | 0.7289E-2 | 0.6202 |
| *y* | 1 | 0.3849 | 0.1312 | 0.5170E-2 |
| *y* | 2 | −0.4805 | 0.1182 | 0.1828E-2 |
| *z* | 1 | 0.7226 | 0.1362 | 0.3004E-5 |
| *z* | 2 | −0.1366E-1 | 0.1336 | 0.9190 |
| Constante | 0 | −0.2387E-2 | 0.8241E-3 | 0.5701E-2 |

$\bar{R}^2 = 0.6503$; $Q(21) = 15.6182$ $(P = 0.7907)$

Pruebas para la significancia conjunta, variable dependiente = *z*

| Variable | Estadístico *F* | Nivel de significancia |
|---|---|---|
| *x* | 0.7396 | 0.4827 |
| *y* | 8.2714 | 0.8360E-3 |
| *z* | 27.9609 | 0.1000E-7 |

* Términos con dos rezagos de cada variable.

## 22.10 Medición de la volatilidad de las series de tiempo financieras: modelos ARCH y GARCH

Como mencionamos en la introducción de este capítulo, las series de tiempo financieras, como precios de acciones, tasas de cambio, tasas de inflación, etc., a menudo presentan el fenómeno de **acumulación de la volatilidad**; es decir, existen lapsos en los que sus precios muestran amplias variaciones durante prolongados periodos y luego se dan intervalos de una calma relativa. Como señala Philip Franses:

> En vista de que tales datos (las series de tiempo financieras) reflejan el resultado del comercio entre compradores y vendedores a precios del mercado, por ejemplo, diversas fuentes de noticias y otros acontecimientos exógenos pueden afectar el patrón de las series de tiempo de los precios de los valores. En vista de que las noticias tienen diversas interpretaciones y también de que los acontecimientos económicos específicos, como las crisis del petróleo, perduran cierto tiempo, a menudo vemos que las grandes observaciones positivas y negativas en las series de tiempo financieras tienden a aparecer en grupos.[21]

Conocer la volatilidad es muy importante en muchas áreas. Por ejemplo, existe una enorme cantidad de trabajo en econometría sobre la variabilidad de la inflación a lo largo del tiempo. Para algunas personas con poder de decisión, la inflación en sí misma quizá no sea dañina, pero no es deseable su variabilidad porque dificulta la planificación financiera.

Sucede lo mismo con los importadores, exportadores y comerciantes que acuden a los mercados de cambio de divisas, pues la variabilidad de las tasas de cambio representa grandes pérdidas o ganancias. A los inversionistas de las casas de bolsa obviamente les interesa la volatilidad de los precios de las acciones, pues una gran volatilidad puede significar enormes pérdidas o ganancias y, en consecuencia, provocar mayor incertidumbre. En los mercados volátiles, a las compañías les resulta difícil capitalizarse en los mercados de capital.

¿Cómo se modelan las series de tiempo financieras que experimentan tal volatilidad? Por ejemplo, ¿cómo se elabora un modelo de las series de tiempo de precios de acciones, tasas de cambio, inflación, etcétera? Una característica de la mayoría de estas series de tiempo financieras consiste en que en su *forma de nivel* son caminatas aleatorias; es decir, son no estacionarias. Por otra parte, en la forma de primeras diferencias, por lo general resultan estacionarias, como vimos en el caso de las series del PIB del capítulo anterior, aunque el PIB no sea estrictamente una serie de tiempo financiera.

En consecuencia, en vez de modelar las series de tiempo financieras en su forma de nivel, ¿por qué no hacer los modelos de sus primeras diferencias? Sin embargo, estas primeras diferencias suelen presentar amplias variaciones, o **volatilidad**, lo cual indica que la varianza de las series de tiempo financieras se modifica con el tiempo. ¿Cómo podemos determinar el modelo de dicha "variación cambiante"? En estos casos es cuando resulta práctico el llamado modelo de **heteroscedasticidad condicional autorregresivo** (ARCH), que originalmente desarrolló Engle.[22]

Como su nombre lo indica, la heteroscedasticidad, o varianza desigual, puede tener una estructura autorregresiva en la que la heteroscedasticidad observada a lo largo de diferentes periodos quizá esté autocorrelacionada. Para tener idea de todo lo que esto significa, analizaremos un ejemplo concreto.

---

[21] Philip Hans Franses, *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Nueva York, 1998, p. 155.

[22] R. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", en *Econometrica*, vol. 50, núm. 1, 1982, pp. 987-1007. Véase también A. Bera y M. Higgins, "ARCH Models: Properties, Estimation and Testing", en *Journal of Economic Surveys*, vol. 7, 1993, pp. 305-366.

**EJEMPLO 22.1**

*Tipo de cambio Estados Unidos/Reino Unido: Ejemplo*

La figura 22.6 proporciona los *logaritmos* del tipo de cambio mensual entre Estados Unidos y el Reino Unido (dólares por libra) de 1971 a 2007, para un total de 444 observaciones mensuales. Como observa en esta figura, hay considerables altibajos en el tipo de cambio a lo largo del periodo muestral. Para ver este movimiento de forma más ilustrativa, en la figura 22.7 graficamos las variaciones en los logaritmos del tipo de cambio; observe que las variaciones en los logaritmos de una variable denotan cambios relativos, los cuales, si se multiplican por 100, proporcionan los cambios porcentuales. Como se observa, los cambios relativos en el tipo de cambio EUA/RU muestran periodos de una amplia variación durante cierto tiempo, y lapsos de fluctuación más bien moderada, lo cual ejemplifica el fenómeno de acumulación de la volatilidad.

Ahora bien, la pregunta práctica es: ¿cómo medimos estadísticamente la volatilidad? Ilustraremos esto con el ejemplo del tipo de cambio.

Sea $Y_t =$ tipo de cambio Estados Unidos/RU

$Y_t^* =$ log de $Y_t$

$dY_t^* = Y_t^* - Y_{t-1}^* =$ variación relativa del tipo de cambio

$d\bar{Y}_t^* =$ media de $dY_t^*$

$X_t = dY_t^* - d\bar{Y}_t^*$

**FIGURA 22.6**
Logaritmo del tipo de cambio Estados Unidos/Reino Unido, 1971-2007 (mensual).

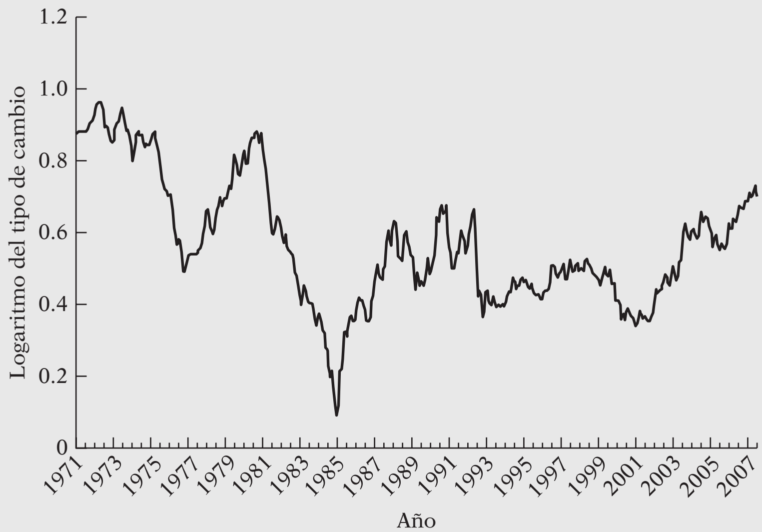
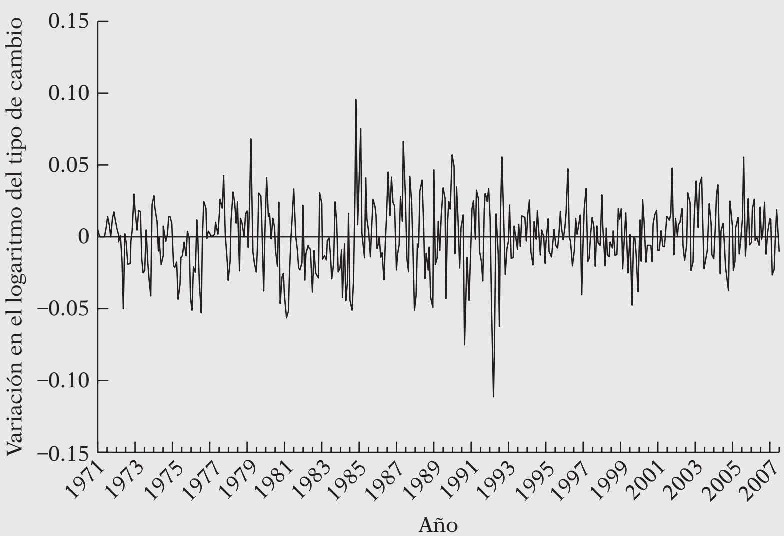



**FIGURA 22.7**
Variación en el logaritmo del tipo de cambio Estados Unidos/Reino Unido.



**EJEMPLO 22.1**
(*continuación*)

Por consiguiente, $X_t$ es el cambio relativo ajustado por la media en el tipo de cambio. Ahora podemos utilizar $X_t^2$ como una medida de la volatilidad. Al ser una cantidad al cuadrado, su valor será alto en periodos en que se experimenten grandes cambios en los precios de los bienes financieros, y comparativamente pequeño cuando sucedan cambios modestos en los precios de dichos bienes.[23]

Al aceptar $X_t^2$ como medida de la volatilidad, ¿cómo sabemos si cambia con el tiempo? Suponga que consideramos el siguiente modelo AR(1) o ARIMA (1, 0, 0):

$$X_t^2 = \beta_0 + \beta_1 X_{t-1}^2 + u_t \tag{22.10.1}$$

Este modelo postula que la volatilidad en el periodo actual está relacionada con su valor del periodo anterior más un término de error de ruido blanco. Si $\beta_1$ es positiva, esto indica que si la volatilidad era alta en el periodo anterior, seguirá siendo alta en el periodo actual, lo cual señala una acumulación de volatilidad. Si $\beta_1$ es cero, no se presenta una acumulación de volatilidad. La significancia estadística del $\beta_2$ estimado se juzga mediante la prueba $t$ usual.

No existe nada que impida considerar un modelo AR($p$) de volatilidad tal que

$$X_t^2 = \beta_0 + \beta_1 X_{t-1}^2 + \beta_2 X_{t-2}^2 + \cdots + \beta_p X_{t-p}^2 + u_t \tag{22.10.2}$$

Este modelo indica que la volatilidad en el periodo actual está relacionada con la de los $p$ periodos anteriores, y el valor de $p$ es una cuestión empírica que se resuelve mediante uno o más criterios de selección de modelos que estudiamos en el capítulo 13 (por ejemplo, la medida de información de Akaike). Podemos probar la significancia de cualquier coeficiente individual $\beta$ mediante la prueba $t$ y la significancia colectiva de dos o más coeficientes mediante la prueba $F$ usual.

El modelo (22.10.1) es un ejemplo de un modelo **ARCH(1)**, y el (22.10.2) se conoce como modelo **ARCH($p$)**, donde $p$ representa el número de términos autorregresivos en el modelo.

Antes de continuar, ejemplificaremos el modelo ARCH con los datos del tipo de cambio entre EUA y RU. Los resultados del modelo ARCH(1) fueron los siguientes:

$$\begin{aligned} X_t^2 &= 0.00043 + 0.23036 X_{t-1}^2 \\ t &= (7.71) \qquad (4.97) \\ & \qquad R^2 = 0.0531 \qquad d = 1.9933 \end{aligned} \tag{22.10.3}$$

donde $X_t^2$ se definió como antes.

Como el coeficiente del término rezagado es en gran medida significativo (el valor $p$ es de casi 0.000), parece que hay acumulación de la volatilidad en este ejemplo. Consideramos modelos ARCH de órdenes más grandes, pero sólo el AR(1) resultó significativo.

En general, ¿cómo probaríamos el efecto ARCH en un modelo de regresión basado en datos de series de tiempo? Para ser más específicos, considere el modelo de regresión con $k$ variables:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \tag{22.10.4}$$

y suponga que al ser *condicional* respecto de la información disponible en el tiempo $(t-1)$, el término de perturbación está distribuido como

$$u_t \sim N\left[0, \left(\alpha_0 + \alpha_1 u_{t-1}^2\right)\right] \tag{22.10.5}$$

(*continúa*)

[23] Tal vez se pregunte por qué no utilizar la varianza de $X_t = \sum X_t^2/n$ como medida de la volatilidad. Lo anterior no es posible, pues deseamos tomar en cuenta la volatilidad cambiante de los precios de las acciones a lo largo del tiempo. Si utilizamos la varianza de $X_t$, sólo constituiría un único valor para el conjunto de datos dado.

**EJEMPLO 22.1** (*continuación*)

es decir, $u_t$ sigue la distribución normal con media 0 y

$$\text{var}(u_t) = \left(\alpha_0 + \alpha_1 u_{t-1}^2\right) \tag{22.10.6}$$

es decir, la varianza de $u_t$ sigue un proceso ARCH(1).

La normalidad de $u_t$ no resulta una novedad. Lo que sí es novedoso es que la varianza de $u$ en el tiempo $t$ dependa de la perturbación al cuadrado en el tiempo $(t-1)$, para dar así la apariencia de una correlación serial.[24] Desde luego, la varianza del error puede depender no sólo del término rezagado del término de error al cuadrado, sino también de diversos términos de error al cuadrado rezagados, de la siguiente forma:

$$\text{var}(u_t) = \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2 \tag{22.10.7}$$

Si no hay autocorrelación en la varianza del error, tenemos

$$H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0 \tag{22.10.8}$$

en cuyo caso var($u_t$) $= \alpha_0$, por lo que tendríamos el efecto ARCH.

Como $\sigma_t^2$ no observamos de manera directa, Engle demostró que al llevar a cabo la siguiente regresión se prueba con facilidad la hipótesis nula anterior:

$$\hat{u}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{u}_{t-1}^2 + \hat{\alpha}_2 \hat{u}_{t-2}^2 + \cdots + \hat{\alpha}_p \hat{u}_{t-p}^2 \tag{22.10.9}$$

donde $\hat{u}_t$, como siempre, denota los residuales de MCO obtenidos del modelo de regresión original (22.10.4).

Se puede demostrar la hipótesis nula $H_0$ mediante la prueba $F$ usual, o de otra manera, al calcular $nR^2$, donde $R^2$ es el coeficiente de determinación obtenido de la regresión auxiliar (22.10.9). Se puede probar que

$$nR^2_{\text{asy}} \sim \chi_p^2 \tag{22.10.10}$$

es decir, para muestras grandes $nR^2$ sigue la distribución ji cuadrada cuyos gl son iguales al número de términos autorregresivos de la regresión auxiliar.

Antes de ejemplificar, hay que asegurarse de no confundir la autocorrelación del término de error, como la analizamos en el capítulo 12, con el modelo ARCH. En dicho modelo, la varianza (condicional) de $u_t$ depende de los términos de error (al cuadrado) anteriores, por lo que da la impresión de que existe autocorrelación.

**EJEMPLO 22.2** *Cambio de precios de las acciones en la Bolsa de Valores de Nueva York*

Como otra ejemplificación del efecto ARCH, en la figura 22.8 presentamos el índice de cambio porcentual mensual de la BVNY (Bolsa de Valores de Nueva York) de 1966 a 2002.[25] Resulta evidente, a partir de la gráfica, que los cambios porcentuales de los precios del índice de la BVNY presentan una volatilidad considerable. Observe sobre todo la enorme variación alrededor de la caída de los precios de las acciones en 1987.

Para expresar la volatilidad en los rendimientos de la bolsa de la figura, considere un modelo muy sencillo:

$$Y_t = \beta_1 + u_t \tag{22.10.11}$$

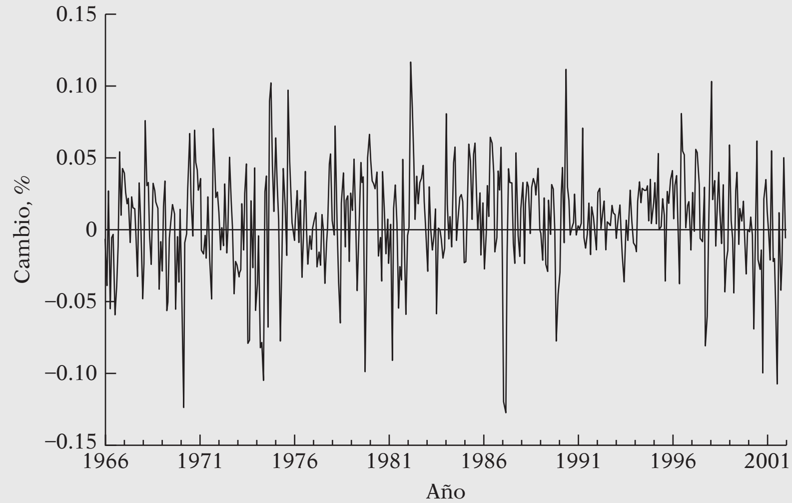
donde $Y_t$ = cambio porcentual en el índice de la BVNY y $u_t$ = término de error aleatorio.

[24] Una nota técnica: Recuerde que, para el modelo clásico lineal, supusimos que la varianza de $u_t$ era igual a $\sigma^2$, que en el presente contexto se convierte en varianza incondicional. Si $\alpha_1 < 1$, la condición de estabilidad, escribimos $\sigma^2 = \alpha_0 + \alpha_1\sigma^2$; es decir, $\sigma^2 = \alpha_0/(1 - \alpha_1)$. Lo anterior demuestra que la varianza incondicional de $u$ no depende de $t$, pero sí del parámetro ARCH $\alpha_1$.

[25] Esta gráfica y los resultados de la regresión presentados después se basan en los datos recopilados por Gary Koop, *Analysis of Economic Data*, John Wiley & Sons, Nueva York, 2000 (los datos se tomaron del disco que acompaña a esta obra). El porcentaje de cambio mensual en el índice de precios de las acciones puede considerarse una tasa de rendimiento del índice.

**EJEMPLO 22.2** (*continuación*)

**FIGURA 22.8** Cambio porcentual mensual del índice de precios de la BVNY, 1966-2002.




Observe que, aparte del intercepto, no existe otra variable explicativa en el modelo. De los datos, obtenemos la siguiente regresión por MCO:

$$\begin{aligned} \hat{Y}_t &= 0.00574 \\ t &= (3.36) \\ d &= 1.4915 \end{aligned} \tag{22.10.12}$$

¿Qué denota este intercepto? Tan sólo la tasa porcentual promedio de rendimiento del índice de la BVNY, o el valor medio de $Y_t$ (¿puede verificar esto?). Así, en el periodo muestral, el rendimiento promedio mensual del índice de la BVNY fue de casi 0.00574%.

Ahora obtenemos los residuos a partir de la regresión anterior y estimamos el modelo ARCH(1), lo cual proporciona los siguientes resultados:

$$\begin{aligned} \widehat{\hat{u}_t^2} &= 0.000007 + 0.25406\hat{u}_{t-1}^2 \\ t &= (0.000) \qquad (5.52) \\ & R^2 = 0.0645 \qquad d = 1.9464 \end{aligned} \tag{22.10.13}$$

donde $\hat{u}_t$ es el residuo estimado de la regresión (22.10.12).

Como el término de perturbación rezagado al cuadrado es estadísticamente significativo (el valor de $p$ es de casi 0.000), parece que las varianzas de los errores están correlacionadas; es decir, existe un efecto ARCH. Probamos modelos ARCH de orden superior, pero sólo ARCH(1) fue estadísticamente significativo.

## ¿Qué hacer cuando ARCH está presente?

Recuerde que ya analizamos diversos métodos para corregir la heteroscedasticidad, lo cual básicamente está implícito al aplicar el método de MCO para transformar los datos. Debe tener presente que MCO aplicados a los datos transformados son mínimos cuadrados generalizados (MCG). Si se encuentra el efecto ARCH, tendremos que utilizar MCG. No daremos detalles técnicos, pues trascienden el alcance de esta obra.[26] Por fortuna, los software como *EViews*, SHAZAM, MICROFIT y PC-GIVE ahora cuentan con rutinas sencillas para estimar dichos modelos.

[26] Consulte Russell Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, Nueva York, 1993. Véase la sec. 16.4 y William H. Greene, *Econometric Analysis*, 4a. ed., Prentice Hall, Englewood Cliffs, Nueva Jersey, 2000, sec. 18.5.

### Advertencia sobre la prueba *d* de Durbin-Watson y el efecto ARCH

Hemos recordado al lector varias veces que un estadístico *d* significativo tal vez no signifique siempre que exista una autocorrelación significativa en los datos disponibles. Con mucha frecuencia, un valor *d* significativo es una indicación de que existen errores de especificación del modelo, los cuales analizamos en el capítulo 13. Ahora tenemos un error de especificación adicional, el debido al efecto ARCH. Por consiguiente, en una regresión de series de tiempo, si se obtiene un valor *d* significativo, debemos probar la existencia del efecto ARCH antes de aceptar el estadístico *d* plenamente. En el ejercicio 22.23 damos un ejemplo.

### Nota sobre el modelo GARCH

Desde su "descubrimiento", en 1982, la elaboración de modelos ARCH se ha convertido en un área floreciente, con todo tipo de variantes respecto del modelo original. Uno muy popular es el **condicional autorregresivo generalizado con heteroscedasticidad**, propuesto por Bollerslev.[27] El modelo GARCH más simple es el GARCH(1,1), que se expresa como:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \quad \textbf{(22.10.14)}$$

el cual enuncia que la varianza condicional de *u* en el tiempo *t* depende no sólo del término de error al cuadrado del periodo anterior [como sucede en ARCH(1)], sino también de su varianza condicional en el periodo anterior. Este modelo puede generalizarse al modelo GARCH($p$, $q$) en el que existen $p$ términos rezagados del término de error al cuadrado y $q$ términos de las varianzas condicionales rezagadas.

No proporcionamos los detalles técnicos de estos modelos, como están presentes; pero sí diremos que el modelo GARCH(1, 1) equivale a un ARCH(2) y el modelo GARCH($p$, $q$) al ARCH($p + q$).[28]

Para los ejemplos del tipo de cambio entre el dólar estadounidense y la libra inglesa, y el de los rendimientos de las acciones en la BVNY, ya establecimos que un modelo ARCH(2) no era significativo, lo cual indica que un modelo GARCH(1, 1) quizá no resulta apropiado en estos casos.

## 22.11 Ejemplos para concluir

Este capítulo finaliza con el estudio de unos cuantos ejemplos adicionales que ilustran algunos puntos que analizamos aquí.

**EJEMPLO 22.3**
*Relación entre el índice de oferta-demanda de empleo (IOD) y la tasa de desempleo (TD) de enero de 1969 a enero de 2000.*

Para analizar la causalidad entre el IOD y la TD, dos indicadores de las condiciones del mercado de trabajo en Estados Unidos, Marc A. Giammatteo planteó el siguiente modelo de regresión:[29]

$$\text{IOD}_t = \alpha_0 + \sum_{i=1}^{25} \alpha_i \text{TD}_{t-i} + \sum_{j}^{25} \beta_j \text{IOD}_{t-j} \quad \textbf{(22.11.1)}$$

$$\text{TD}_t = \alpha_0 + \sum_{i=1}^{25} \lambda_i \text{TD}_{t-i} + \sum_{j=1}^{25} \delta_j \text{IOD}_{t-j} \quad \textbf{(22.11.2)}$$

Con el objeto de ahorrar espacio, no presentamos los resultados propiamente dichos de la regresión, pero la conclusión principal que surge de este análisis es que existe una causalidad bilateral entre los dos indicadores del mercado de trabajo; asimismo, dicha conclusión no cambió cuando se modificó la duración del rezago. Los datos sobre el IOD y la TD se proporcionan en el sitio Web del libro, en la tabla 22.5.

[27] T. Bollerslev, "Generalized Autorregresive Conditional Heteroscedasticity", en *Journal of Econometrics*, vol. 31, 1986, pp. 307-326.

[28] Para mayores detalles, véase Davidson y MacKinnon, *op. cit.*, pp. 558-560.

[29] Marc A. Giammatteo (West Point, Generación de 2000), "The Relationship between the Help-Wanted Index and the Unemployment Rate", trabajo final del trimestre escolar, inédito. (La notación se modificó con el objeto de adaptarla a la de este libro.)

**EJEMPLO 22.4**
*Creación de un modelo ARIMA para el tipo de cambio entre el yen y el dólar: de enero de 1971 a abril de 2008*

El tipo de cambio yen/dólar (¥/$) es un tipo de cambio clave. Al tomar el logaritmo de ¥/$ mensual se encontró que, en la forma nivel, este tipo de cambio mostró el patrón habitual de una serie de tiempo no estacionaria. Pero al analizar las primeras diferencias, se descubrió que eran estacionarias; la gráfica de esta situación se parece mucho a la figura 22.8.

El análisis de raíz unitaria confirmó que las primeras diferencias de los logaritmos de ¥/$ eran estacionarias. Después de examinar el correlograma del logaritmo de las primeras diferencias, estimamos el siguiente modelo MA(1):

$$\begin{aligned} \hat{Y}_t &= -0.0028 - 0.3300u_{t-1} \\ t &= (-1.71) \quad (-7.32) \\ & \qquad R^2 = 0.1012 \qquad d = 1.9808 \end{aligned} \tag{22.11.3}$$

donde $Y_t$ = las primeras diferencias de los logaritmos de ¥/$ y $u$ es un término de error de ruido blanco.

Para ahorrar espacio, proporcionamos los datos en los que se basó el análisis anterior en el sitio Web del libro, en la tabla 22.6. Con esa información, se pide al lector que pruebe otros modelos y compare los pronósticos obtenidos.

**EJEMPLO 22.5**
*Modelo ARCH para la tasa de inflación de Estados Unidos: De enero de 1947 a marzo de 2008*

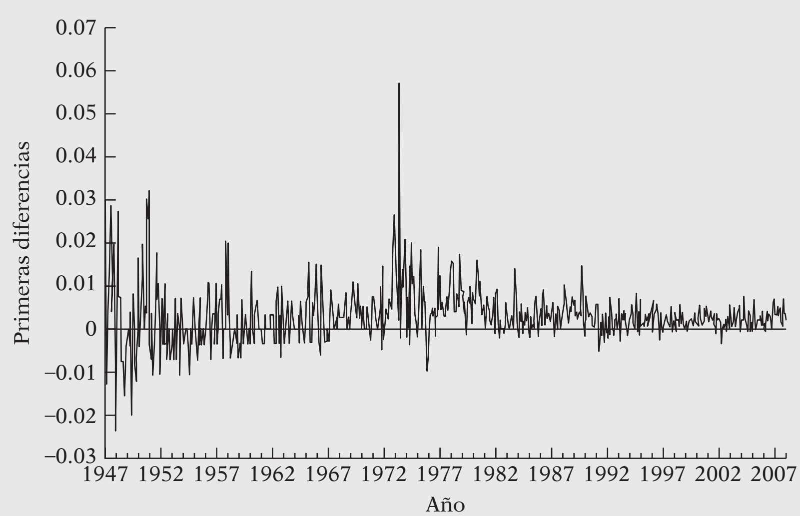
Para verificar si el efecto ARCH está presente en la tasa de inflación de Estados Unidos según la mide el IPC, obtuvimos los datos de dicho índice de enero de 1947 a marzo de 2008. La gráfica de los logaritmos del IPC mostró que la serie de tiempo era no estacionaria. Pero la gráfica de las primeras diferencias de los logaritmos del IPC, como se muestra en la figura 22.9, presenta volatilidad considerable, si bien las primeras diferencias son estacionarias.

Según el procedimiento descrito en las regresiones (22.10.12) y (22.10.13), primero efectuamos la regresión de los logaritmos de las primeras diferencias del IPC sobre una constante y obtenemos los residuos de esta ecuación. Al elevar al cuadrado los residuos tenemos el siguiente modelo ARCH(2):

$$\begin{aligned} \widehat{\hat{u}_t^2} &= 0.000028 + 0.12125\hat{u}_{t-1}^2 + 0.08718\hat{u}_{t-2}^2 \\ t &= (5.42) \qquad (3.34) \qquad (2.41) \\ & \qquad R^2 = 0.026 \qquad d = 2.0214 \end{aligned} \tag{22.11.4}$$

Como puede observar, existe una enorme persistencia de la volatilidad, pues, en el mes actual, depende de la de los últimos tres meses. Se recomienda al lector que obtenga los datos sobre el IPC en fuentes gubernamentales e intente averiguar si otro modelo resulta más adecuado, de preferencia un modelo GARCH.

**FIGURA 22.9**
Primeras diferencias de los logaritmos del IPC.



## Resumen y conclusiones

1. Los métodos Box-Jenkins y VAR para pronósticos económicos son opciones respecto de los modelos tradicionales uniecuacionales y de ecuaciones simultáneas.
2. Para pronosticar valores de una serie de tiempo, la estrategia básica de Box-Jenkins es la siguiente:
   *a*) Examine primero si la serie es estacionaria. Esto se logra al calcular la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP), o mediante un análisis formal de raíz unitaria. Los correlogramas asociados a FAC y FACP son, con frecuencia, buenas herramientas de diagnóstico visual.
   *b*) Si la serie de tiempo es no estacionaria, debe diferenciarse una o más veces para alcanzar la estacionariedad.
   *c*) Se calculan entonces la FAC y la FACP de la serie de tiempo estacionaria para determinar si la serie es autorregresiva pura, del tipo de promedios móviles puro, o una mezcla de las dos. Así, de los lineamientos generales de la tabla 22.1, se determinan los valores de $p$ y $q$ en el proceso ARMA que se va a ajustar. En esta etapa, el modelo ARMA($p$, $q$) seleccionado es tentativo.
   *d*) Entonces se estima el modelo tentativo.
   *e*) Se examinan los residuos de este modelo tentativo para establecer si son de ruido blanco. Si lo son, el modelo tentativo es quizás una buena aproximación al proceso estocástico subyacente. Si no lo son, el proceso se inicia de nuevo. Por consiguiente, el método de Box-Jenkins es iterativo.
   *f*) El modelo finalmente seleccionado sirve para pronosticar.
3. El enfoque VAR para pronósticos considera diferentes series de tiempo a la vez. Las características que distinguen al VAR son las siguientes:
   *a*) Es un sistema simultáneo en el sentido de que todas las variables se consideran endógenas.
   *b*) En la elaboración de modelos VAR, el valor de una variable se expresa como función lineal de los valores pasados o rezagados de esa variable y de todas las demás variables incluidas en el modelo.
   *c*) Si cada ecuación contiene el mismo número de variables rezagadas en el sistema, éste se estima mediante MCO sin recurrir a otros métodos asociados a sistemas, como el método de mínimos cuadrados en dos etapas (MC2E) o las regresiones aparentemente no relacio*nadas (SURE).*
   *d*) Esta simplicidad en la elaboración de modelos VAR puede ser su desventaja. En vista del limitado número de observaciones disponibles en la mayoría de los análisis económicos, la introducción de los diversos rezagos de cada variable puede consumir muchos grados de libertad.[30]
   *e*) Si hay varios rezagos en cada ecuación, no siempre es fácil interpretar cada coeficiente, en especial si se alternan los signos de los coeficientes. Por esta razón, en la elaboración de modelos VAR se examina la función de impulso-respuesta (FIR) para determinar la forma como responde la variable dependiente ante un choque administrado a una o más ecuaciones en el sistema.
   *f*) Hay mucha discusión y controversia sobre la superioridad de los diversos métodos de pronósticos. Los métodos de pronósticos uniecuacionales, de ecuaciones simultáneas, Box-Jenkins y VAR tienen sus defensores como también sus detractores. Todo lo que podemos decir es que no hay un método que por sí mismo sea apropiado para todas las situaciones. De ser así, no habría necesidad de analizar las diversas opciones. Una cosa es segura: las metodologías de Box-Jenkins y VAR ya forman parte integral de la econometría.
4. En este capítulo también estudiamos una clase especial de modelos, ARCH y GARCH, particularmente útiles en el análisis de las series de tiempo financieras, como precios de acciones,

[30] Los partidarios de las estadísticas bayesianas creen que este problema se puede reducir. Véase R. Litterman, "A Statistical Approach to Economic Forecasting", en *Journal of Business and Economic Statistics*, vol. 4, 1986, pp. 1-4.

tasas de inflación y tasas de cambio. Una característica sobresaliente de estos modelos es que la varianza del error quizá esté correlacionada a lo largo del tiempo, debido al fenómeno de la acumulación de la volatilidad. Al respecto, también señalamos que, en muchos casos, una *d* de Durbin-Watson significativa puede deberse a un efecto ARCH o GARCH.

5. Existen variantes de los modelos ARCH y GARCH, pero no los consideramos en este capítulo debido a restricciones de espacio. Algunos de estos otros modelos son **GARCH-M** (GARCH en la media), **TGARCH** (umbral GARCH) y **EGARCH** (GARCH exponencial). En las referencias hay debates sobre estos modelos.[31]

## EJERCICIOS

### Preguntas

22.1. ¿Cuáles son los métodos más importantes para pronósticos económicos?

22.2. ¿Cuáles son las principales diferencias entre el método de ecuaciones simultáneas y el de Box-Jenkins para pronósticos económicos?

22.3. Esquematice los pasos principales relacionados con la aplicación del método de Box-Jenkins para pronósticos económicos.

22.4. ¿Qué sucede si se aplican las técnicas de Box-Jenkins a series de tiempo no estacionarias?

22.5. ¿Qué diferencias hay entre los métodos de Box-Jenkins y VAR para pronósticos económicos?

22.6. ¿En qué sentido es ateórico el modelo VAR?

22.7. "Si el objetivo principal es el pronóstico, VAR lo logrará." Evalúe críticamente esta afirmación.

22.8. Como el número de rezagos que se va a introducir en un modelo VAR puede ser un asunto subjetivo, ¿cómo se decide cuántos rezagos deben introducirse en una aplicación concreta?

22.9. Comente la siguiente afirmación: "Box-Jenkins y VAR son los ejemplos más importantes de medición sin teoría".

22.10. ¿Cuál es la conexión, de existir, entre las pruebas de causalidad de Granger y el diseño de modelos VAR?

### Ejercicios empíricos

22.11. Considere los datos sobre los logaritmos del IPD (ingreso personal disponible) en la sección 21.1 (consulte los datos reales en el sitio Web del libro). Suponga que deseamos ajustar un modelo ARIMA apropiado a estos datos. Defina los pasos que implica la realización de esta labor.

22.12. Repita el ejercicio 22.11 para los datos de LGCP (gasto de consumo personal) presentados en la sección 21.1 (una vez más, consulte los datos reales en el sitio Web del libro).

22.13. Repita el ejercicio 22.11 para los datos de LUE (utilidades empresariales).

22.14. Repita el ejercicio 22.11 para los datos de LDIVIDENDOS.

22.15. En la sección 13.9 explicamos el criterio de información de Schwarz para determinar la longitud del rezago. ¿Cómo utilizaría este criterio para determinar la longitud apropiada del rezago en un modelo VAR?

22.16. Con los datos sobre LGCP y LIPD de la sección 21.1 (consulte los datos reales en el sitio Web del libro), desarrolle un modelo bivariado VAR de I-1970 a IV-2006. Con este modelo pronostique los valores de estas variables para los cuatro trimestres de 2007 y compare los valores de pronóstico con los valores reales que se proporcionan en el conjunto de datos.

[31] Véase Walter Enders, *Applied Econometric Time Series,* 2a. ed., John Wiley & Sons, Nueva York, 2004. Para un análisis orientado a la aplicación, véase Dimitrios Asteriou y Stephen Hall, *Applied Econometrics: A Modern Approach,* edición revisada, Palgrave/Macmillan, Nueva York, 2007, capítulo 14.

22.17. Repita el ejercicio 22.16 con los datos sobre LDIVIDENDOS y LUE.

*22.18. Utilice cualquier paquete estadístico y estime la función de impulso-respuesta para un periodo hasta de ocho rezagos para el modelo VAR que desarrolló en el ejercicio 22.16.

22.19. Repita el ejercicio 22.18 para el modelo VAR desarrollado en el ejercicio 22.17.

22.20. Consulte los resultados de la regresión VAR de la tabla 22.4. De las diversas pruebas $F$ hechas en las tres regresiones que allí se dieron, ¿qué puede decir sobre la naturaleza de la causalidad en las tres variables?

22.21. Continuando con el ejercicio 20.20, ¿puede suponer la razón por la cual los autores decidieron expresar las tres variables en el modelo en forma de cambios porcentuales en lugar de usar las variables en su forma de nivel? (*Sugerencia:* Estacionariedad.)

22.22. Con los datos para Canadá proporcionados en la tabla 17.5, investigue si $M_1$ y $R$ son variables aleatorias estacionarias. Si no es así, ¿están cointegradas? Muestre los cálculos necesarios.

22.23. Continúe con los datos de la tabla 17.5. Ahora considere el siguiente modelo sencillo de la demanda de dinero en Canadá:

$$\ln M_{1t} = \beta_1 + \beta_2 \ln \text{PIB}_t + \beta_3 \ln R_t + u_t$$

*a*) ¿Cómo interpretaría los parámetros de este modelo?

*b*) Obtenga los residuos a partir de este modelo y determine si existe algún efecto ARCH.

22.24. Consulte el modelo ARCH(3) dado en (22.11.4). Con los mismos datos calculamos el siguiente modelo ARCH(1):

$$\begin{aligned} \widehat{\hat{u}_t^2} &= 0.00000078 + 0.3737\hat{u}_{t-1}^2 \\ t &= (7.5843) \qquad\qquad (10.2351) \\ & \qquad\qquad R^2 = 0.1397 \qquad d = 1.9896 \end{aligned}$$

¿Cómo elegiría entre ambos modelos? Muestre los cálculos necesarios.

22.25. La tabla 22.7 contiene datos sobre las tasas de interés de los pagarés del Tesoro a tres (TB3M) y a seis meses (TB6M), del 1 de enero de 1982 a marzo de 2008, para un total de 315 observaciones mensuales. Los datos se encuentran en el sitio Web del libro.

*a*) Trace la gráfica de las dos series de tiempo en el mismo diagrama. ¿Qué observa?

*b*) Realice un análisis formal de raíz unitaria para ver si estas series de tiempo son estacionarias.

*c*) ¿Están cointegradas las dos series de tiempo? ¿Cómo sabe? Realice los cálculos necesarios.

*d*) ¿Qué significado económico tiene *cointegración* en este contexto? Si las dos series no están cointegradas, ¿qué repercusiones económicas tiene esto?

*e*) Si desea estimar un modelo VAR, por ejemplo, con cuatro rezagos de cada variable, ¿es necesario usar las primeras diferencias de las dos series o puede realizar el análisis de las dos series en su forma de nivel? Justifique su respuesta.

22.26. *Ejercicio en clase:* Elija un índice del mercado de valores y obtenga datos *diarios* sobre el valor del índice elegido para cinco años con el propósito de averiguar si el índice bursátil se caracteriza por efectos ARCH.

22.27. *Ejercicio en clase:* Recopile datos sobre las tasas de inflación y desempleo en Estados Unidos correspondientes a los periodos trimestrales de 1980 a 2007 y desarrolle y estime un modelo VAR para las dos variables. Para calcular la tasa de inflación, utilice el IPC (índice de precios al consumidor), y la tasa de desempleo civil para calcular la tasa de desempleo. Es preciso prestar mucha atención a la estacionariedad de estas variables. Además, averigüe si una variable causa, según Granger, la otra variable. Presente todos sus cálculos.

---

* Opcional.

# Apéndice A

# Revisión de algunos conceptos estadísticos

En este apéndice se introducen, en forma muy general, algunos conceptos estadísticos que aparecen en este texto. El análisis no es riguroso y no se presentan pruebas debido a que existen diversos libros de estadística, excelentes, que hacen muy bien ese trabajo. Algunos de esos libros se mencionan al final del apéndice.

## A.1 Operadores de sumatoria y de producto

Con la letra mayúscula griega $\sum$ (sigma) se indica la sumatoria. Así,

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

Algunas de las propiedades más importantes del operador de sumatoria $\sum$ son

1. $\sum_{i=1}^{n} k = nk$, donde $k$ es una constante. Así, $\sum_{i=1}^{4} 3 = 4 \cdot 3 = 12$.
2. $\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i$, donde $k$ es una constante.
3. $\sum_{i=1}^{n}(a + bx_i) = na + b\sum_{i=1}^{n} x_i$, donde $a$ y $b$ son constantes y se emplean las propiedades 1 y 2 anteriores.
4. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$.

El operador de sumatoria también se amplía a sumas múltiples. Así, $\sum\sum$, el operador de doble sumatoria, se define como

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} &= \sum_{i=1}^{n}(x_{i1} + x_{i2} + \cdots + x_{im}) \\ &= (x_{11} + x_{21} + \cdots + x_{n1}) + (x_{12} + x_{22} + \cdots + x_{n2}) \\ &\quad + \cdots + (x_{1m} + x_{2m} + \cdots + x_{nm})\end{aligned}$$

Algunas de las propiedades de $\sum\sum$ son:

1. $\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij}$; es decir, el orden en el cual se realice la doble sumatoria es intercambiable.
2. $\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} y_j$.

3. $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}+y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}$.
4. $\left[\sum_{i=1}^{n}x_i\right]^2 = \sum_{i=1}^{n}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_ix_j = \sum_{i=1}^{n}x_i^2 + 2\sum_{i<j}x_ix_j$.

El operador de producto $\Pi$ se define como

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdots x_n$$

Por tanto,

$$\prod_{i=1}^{3} x_i = x_1 \cdot x_2 \cdot x_3$$

## A.2 Espacio muestral, puntos muestrales y sucesos

El conjunto de todos los resultados posibles de un experimento aleatorio, o al azar, se denomina **población** o **espacio muestral**, y cada miembro de este espacio muestral se denomina **punto muestral**. Por tanto, en el experimento de lanzar dos monedas, el espacio muestral consta de estos cuatro resultados posibles: *HH, HT, TH* y *TT*, donde *HH* significa una cara en el primer lanzamiento y nuevamente una cara en el segundo, *HT* significa una cara en el primer lanzamiento y una cruz en el segundo, y así sucesivamente. Cada uno de los sucesos anteriores constituye un punto muestral.

Un **suceso** es un subconjunto del espacio muestral. Así, si *A* denota la ocurrencia de una cara y de una cruz, entonces, de los posibles resultados anteriores, sólo dos pertenecen a *A*, a saber: *HT* y *TH*. En este caso, *A* constituye un suceso. En forma similar, la ocurrencia de dos caras en el lanzamiento de dos monedas es un suceso. Se dice que los sucesos son **mutuamente excluyentes** si la ocurrencia de uno impide la ocurrencia de otro. Si en el ejemplo anterior ocurre *HH*, no es posible la ocurrencia del suceso *HT* al mismo tiempo. Decimos que los sucesos son **exhaustivos** (colectivamente) si se agotan todos los resultados posibles de un experimento. Así, en el ejemplo, los sucesos *a*) dos caras, *b*) dos cruces y *c*) una cruz y una cara agotan todos los resultados posibles; por tanto, son sucesos exhaustivos (colectivamente).

## A.3 Probabilidad y variables aleatorias

### Probabilidad

Sea *A* un suceso en un espacio muestral. Sea $P(A)$ la probabilidad del suceso *A*, es decir, la proporción de veces que el suceso *A* ocurrirá en ensayos repetidos de un experimento. En forma alterna, en un total de *n* posibles resultados igualmente probables de un experimento, si *m* de ellos son favorables a la ocurrencia del suceso *A*, se define la razón *m/n* como la **frecuencia relativa** de *A*. Para valores grandes de *n*, esta frecuencia relativa constituye una muy buena aproximación de la probabilidad de *A*.

#### *Propiedades de la probabilidad*

$P(A)$ es una función de valor real[1] y tiene estas propiedades:

1. $0 \leq P(A) \leq 1$ para todo *A*.
2. Si $A, B, C, \ldots$ constituye un conjunto de sucesos exhaustivo, entonces $P(A+B+C+\cdots) = 1$, donde $A+B+C$ significa *A* o *B* o *C*, y así sucesivamente.
3. Si $A, B, C, \ldots$ son sucesos mutuamente excluyentes, entonces

$$P(A+B+C+\cdots) = P(A) + P(B) + P(C) + \cdots$$

[1] Una función cuyo dominio y rango son subconjuntos de números reales se conoce generalmente como función real de una variable real. Para mayores detalles, véase Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, 3a. ed., McGraw-Hill, 1984, capítulo 2.

**EJEMPLO 1**

Considere el experimento de lanzar un dado numerado del 1 al 6. El espacio muestral consta de los resultados 1, 2, 3, 4, 5 y 6. Por consiguiente, estos seis sucesos agotan la totalidad del espacio muestral. La probabilidad de obtener cualquiera de estos números es 1/6, pues son seis resultados igualmente probables y cada uno tiene igual probabilidad de aparecer. Como 1, 2, 3, 4, 5 y 6 forman un conjunto exhaustivo de sucesos, $P(1 + 2 + 3 + 4 + 5 + 6) = 1$, donde 1, 2, 3,... significa la probabilidad del número 1 o del número 2 o del número 3, etcétera. Como 1, 2,..., 6 son sucesos mutuamente excluyentes en donde dos números no pueden obtenerse simultáneamente, $P(1 + 2 + 3 + 4 + 5 + 6) = P(1) + P(2) + \cdots + P(6) = 1$.

### Variables aleatorias

Una variable cuyo valor está determinado por el resultado de un experimento al azar se denomina **variable aleatoria** (va). Las variables aleatorias se denotan usualmente por las letras mayúsculas $X$, $Y$, $Z$, y así sucesivamente, y los valores que toman se denotan por letras minúsculas, $x$, $y$, $z$, etcétera.

Una variable aleatoria puede ser **discreta** o **continua**. Una va discreta adquiere sólo un número finito (o infinito contable) de valores.[2] Por ejemplo, al lanzar dos dados, cada uno numerado del 1 al 6, si definimos la variable aleatoria $X$ como la suma de los números que aparecen en los dados, entonces $X$ toma uno de los siguientes valores: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12; por tanto, se trata de una variable aleatoria discreta. Una va continua, por su parte, es una variable que puede tomar cualquier valor dentro de un intervalo de valores. Así, la estatura de un individuo es una variable continua —por ejemplo, en el intervalo entre 152.4 y 165.1 centímetros— y puede adquirir cualquier valor, según la precisión de la medición.

## A.4 Función de densidad de probabilidad (FDP)

### Función de densidad de probabilidad de una variable aleatoria discreta

Sea $X$ una va discreta que toma valores diferentes $x_1, x_2, \ldots x_n$. Entonces, la función

$$\begin{aligned} f(x) &= P(X = x_i) \qquad \text{para } i = 1, 2, \ldots, n, \ldots \\ &= 0 \qquad\qquad\quad\ \ \text{para } x \neq x_i \end{aligned}$$

se denomina **función de densidad de probabilidad discreta** (FDP) de $X$, donde $P(X = x_i)$ significa la probabilidad de que la va discreta $X$ tome el valor de $x_i$.

**EJEMPLO 2**

En un lanzamiento de dos dados la variable aleatoria $X$, la suma de los números que aparecen en dos dados, puede tomar uno de los 11 valores mostrados. La FDP de esta variable se muestra como sigue (véase también la figura A.1):

$$\begin{array}{lccccccccccc} x = & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ f(x) = & \left(\frac{1}{36}\right) & \left(\frac{2}{36}\right) & \left(\frac{3}{36}\right) & \left(\frac{4}{36}\right) & \left(\frac{5}{36}\right) & \left(\frac{6}{36}\right) & \left(\frac{5}{36}\right) & \left(\frac{4}{36}\right) & \left(\frac{3}{36}\right) & \left(\frac{2}{36}\right) & \left(\frac{1}{36}\right) \end{array}$$

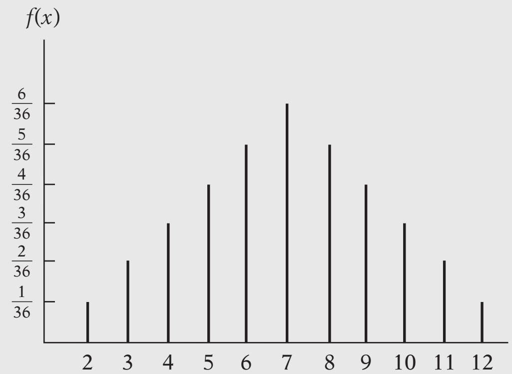
Estas probabilidades se verifican fácilmente. En total, hay 36 resultados posibles, de los cuales uno es favorable al número 2, dos son favorables al número 3 (pues la suma de 3 se presenta como 1 en el primer dado y 2 en el segundo dado, o 2 en el primer dado y 1 en el segundo dado), y así sucesivamente.

(*continúa*)

[2] Hay un análisis sencillo de la noción de conjuntos contables infinitos en R. G. D. Allen, *Basic Mathematics*, Macmillan, Londres, 1964, p. 104.

**EJEMPLO 2** (*continuación*)

**FIGURA A.1** Función de densidad de la variable aleatoria discreta del ejemplo 2.




## Función de densidad de probabilidad de una variable aleatoria continua

Sea $X$ una va continua. Entonces, se dice que $f(x)$ es la FDP de $X$ si se satisfacen las siguientes condiciones:

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{a}^{b} f(x)\,dx = P(a \leq x \leq b)$$

donde $f(x)dx$ se conoce como el elemento probabilístico (la probabilidad asociada a un pequeño intervalo de una variable continua) y donde $P(a \leq X \leq b)$ significa la probabilidad de que $X$ se encuentre en el intervalo $a$ a $b$. Geométricamente, tenemos la figura A.2.

Para una va continua, en contraste con una va discreta, la probabilidad de que $X$ tome un valor específico es cero;[3] la probabilidad para tal variable sólo se mide sobre un rango o intervalo dado, como $(a, b)$ que aparece en la figura A.2.

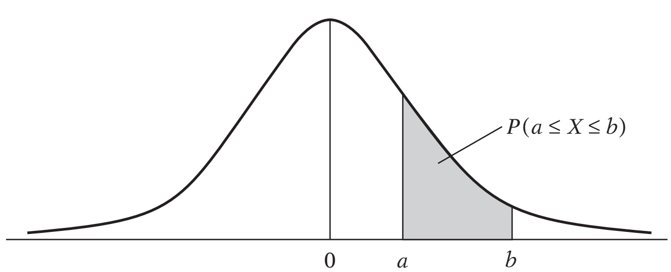
**EJEMPLO 3**

Considere la siguiente función de densidad:

$$f(x) = \frac{1}{9}x^2 \qquad 0 \leq x \leq 3$$

Se verifica con facilidad que $f(x) \geq 0$ para toda $X$ en el rango 0 a 3 y que $\int_0^3 \frac{1}{9}x^2dx = 1$. (*Nota:* La integral es $(\frac{1}{27}x^3 |_0^3) = 1$.) Si deseamos evaluar la FDP anterior entre 0 y 1, obtenemos $\int_0^1 \frac{1}{9}x^2dx = (\frac{1}{27}x^3 |_0^1) = \frac{1}{27}$; es decir, la probabilidad de que $X$ se encuentre entre 0 y 1 es de 1/27.

**FIGURA A.2** Función de densidad de una variable aleatoria continua.




[3] *Nota:* $\int_a^a f(x)\,dx = 0$.

## Funciones de densidad de probabilidad conjunta

*FDP conjunta discreta*

Sean $X$ y $Y$ dos variables aleatorias discretas. Entonces, la función

$$\begin{aligned} f(x, y) &= P(X = x \text{ y } Y = y) \\ &= 0 \qquad \text{cuando } X \neq x \text{ y } Y \neq y \end{aligned}$$

se conoce como **función de densidad de probabilidad conjunta discreta** y da la probabilidad (conjunta) de que $X$ tome el valor de $x$ y $Y$ tome el valor de $y$.

**EJEMPLO 4**

La siguiente tabla presenta la FDP conjunta de las variables discretas $X$ y $Y$.

| | | X | | | |
|---|---|---|---|---|---|
| | | **−2** | **0** | **2** | **3** |
| Y | **3** | 0.27 | 0.08 | 0.16 | 0 |
| | **6** | 0 | 0.04 | 0.10 | 0.35 |

Esta tabla muestra que la probabilidad de que $X$ tome el valor de $-2$ mientras simultáneamente $Y$ toma el valor de 3 es 0.27, y que la probabilidad de que $X$ tome el valor de 3 mientras $Y$ toma el valor de 6 es 0.35, y así sucesivamente.

## Función de densidad de probabilidad marginal

En relación con $f(x, y)$, $f(x)$ y $f(y)$ se denominan funciones de densidad de probabilidad **individuales** o **marginales**. Estas FDP marginales se obtienen de la siguiente manera:

$$f(x) = \sum_y f(x, y) \qquad \text{FDP marginal de } X$$

$$f(y) = \sum_x f(x, y) \qquad \text{FDP marginal de } Y$$

donde, por ejemplo, $\sum_y$ significa la suma sobre todos los valores de $Y$ y $\sum_x$ significa la suma sobre todos los valores de $X$.

**EJEMPLO 5**

Considere la información del ejemplo 4. La FDP marginal de $X$ se obtiene de la siguiente manera:

$$f(x = -2) = \sum_y f(x, y) = 0.27 + 0 = 0.27$$

$$f(x = 0) = \sum_y f(x, y) = 0.08 + 0.04 = 0.12$$

$$f(x = 2) = \sum_y f(x, y) = 0.16 + 0.10 = 0.26$$

$$f(x = 3) = \sum_y f(x, y) = 0 + 0.35 = 0.35$$

Asimismo, la FDP marginal de $Y$ se obtiene así:

$$f(y = 3) = \sum_x f(x, y) = 0.27 + 0.08 + 0.16 + 0 = 0.51$$

$$f(y = 6) = \sum_x f(x, y) = 0 + 0.04 + 0.10 + 0.35 = 0.49$$

Como muestra este ejemplo, para obtener la FDP marginal de $X$ sumamos la columna de números y para obtener la FDP marginal de $Y$ sumamos la fila de números. Observe que $\sum_x f(x)$ sobre todos los valores de $X$ es 1, como lo es $\sum_y f(y)$ sobre todos los valores de $Y$ (¿por qué?).

### *FDP condicional*

Como mencionamos en el capítulo 2, en el análisis de regresión, el interés con frecuencia es estudiar el comportamiento de una variable condicional respecto de los valores de otra u otras variables. Para esto se considera la FDP condicional. La función

$$f(x \mid y) = P(X = x \mid Y = y)$$

se conoce como **FDP condicional** de $X$; da la probabilidad de que $X$ tome el valor de $x$ porque $Y$ asumió el valor de $y$. En forma similar,

$$f(y \mid x) = P(Y = y \mid X = x)$$

lo cual da la *FDP condicional de Y*.

Las FDP condicionales se obtienen de la siguiente manera:

$$f(x \mid y) = \frac{f(x, y)}{f(y)} \qquad \text{FDP condicional de } X$$

$$f(y \mid x) = \frac{f(x, y)}{f(x)} \qquad \text{FDP condicional de } Y$$

Como muestran las expresiones anteriores, la FDP condicional de una variable se expresa como la razón de la FDP conjunta respecto de la FDP marginal de otra variable (condicionante).

**EJEMPLO 6**

Continuamos con los ejemplos 4 y 5 y calculamos las siguientes probabilidades condicionales:

$$f(X = -2 \mid Y = 3) = \frac{f(X = -2, Y = 3)}{f(Y = 3)} = 0.27/0.51 = 0.53$$

Observe que la probabilidad incondicional $f(X = -2)$ es 0.27, pero si $Y$ asume el valor de 3, la probabilidad de que $X$ tome el valor de $-2$ es 0.53.

$$f(X = 2 \mid Y = 6) = \frac{f(X = 2, Y = 6)}{f(Y = 6)} = 0.10/0.49 = 0.20$$

Observe de nuevo que la probabilidad incondicional de que $X$ tome el valor de 2 es 0.26, la cual es diferente de 0.20, que es su valor si $Y$ asume el valor de 6.

## Independencia estadística

Dos variables aleatorias $X$ y $Y$ son estadísticamente independientes si y sólo si

$$f(x, y) = f(x)f(y)$$

es decir, si la FDP conjunta se expresa como el producto de las FDP marginales.

**EJEMPLO 7**

Una bolsa contiene tres bolas numeradas 1, 2 y 3. Se seleccionan de la bolsa dos bolas al azar, con reemplazo (es decir, la primera bola sacada se vuelve a poner en la bolsa antes de sacar la segunda). Sea $X$ el número de la primera bola sacada y $Y$ el número de la segunda. La siguiente tabla da la FDP conjunta *de X* y *Y*.

**EJEMPLO 7**
(*continuación*)

| | | X | | |
|---|---|---|---|---|
| | | **1** | **2** | **3** |
| | **1** | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |
| Y | **2** | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |
| | **3** | $\frac{1}{9}$ | $\frac{1}{9}$ | $\frac{1}{9}$ |

Ahora $f(X=1, Y=1) = \frac{1}{9}$, $f(X=1) = \frac{1}{3}$ (obtenido mediante la suma de los elementos de la primera columna) y $f(y = 1) = \frac{1}{3}$ (obtenido mediante la suma de los elementos de la primera fila). Como en este ejemplo $f(X,Y) = f(X)f(Y)$, podemos decir que las dos variables son estadísticamente independientes. Con facilidad se verifica que, para cualquier otra combinación de los valores $X$ y $Y$ de la tabla anterior, las FDP conjuntas se factorizan en FDP individuales.

Se observa que las variables $X$ y $Y$ del ejemplo 4 no son estadísticamente independientes, pues el producto de las dos FDP marginales no es igual a la FDP conjunta. (*Nota:* Debe cumplirse que $f(X, Y) = f(X)f(Y)$ en todas las combinaciones de $X$ y $Y$ si las dos variables han de ser estadísticamente independientes.)

### *FDP conjunta continua*

La FDP $f(x, y)$ de dos variables continuas $X$ y $Y$ es tal que

$$f(x, y) \geq 0$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$$

$$\int_{c}^{d}\int_{a}^{b} f(x, y)\,dx\,dy = P(a \leq x \leq b, c \leq y \leq d)$$

**EJEMPLO 8**

Considere la siguiente FDP:

$$f(x, y) = 2 - x - y \qquad 0 \leq x \leq 1; 0 \leq y \leq 1$$

Es obvio que $f(x, y) \geq 0$. Además,[4]

$$\int_0^1\int_0^1 (2 - x - y)\,dx\,dy = 1$$

La FDP marginal de $X$ y de $Y$ se obtiene como

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \qquad \text{FDP marginal de } X$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \qquad \text{FDP marginal de } Y$$

---

4

$$\int_0^1\left[\int_0^1 (2 - x - y)\,dx\right]dy = \int_0^1\left[\left(2x - \frac{x^2}{2} - xy\right)\Big|_0^1\right]dy$$
$$= \int_0^1\left(\frac{3}{2} - y\right)dy$$
$$= \left(\frac{3}{2}y - \frac{y^2}{2}\right)\Big|_0^1 = 1$$

*Nota:* La expresión $(\frac{3}{2}y - y^2/2)|_0^1$ significa que la expresión entre paréntesis debe evaluarse para el valor del límite superior 1 y para el valor del límite inferior 0; el último valor se resta del primero para obtener el valor de la integral. Así, en el ejemplo anterior, los límites son $(\frac{3}{2} - \frac{1}{2})$ en $y = 1$ y 0 en $y = 0$, de lo que resulta que el valor de la integral es igual a 1.

**EJEMPLO 9** Las dos FDP marginales de la FDP conjunta del ejemplo 8 son las siguientes:

$$f(x) = \int_0^1 f(x, y)dy = \int_0^1 (2 - x - y)dy$$

$$\left(2y - xy - \frac{y^2}{2}\right)\Bigg|_0^1 = \frac{3}{2} - x \qquad 0 \le x \le 1$$

$$f(y) = \int_0^1 (2 - x - y)dx$$

$$\left(2x - xy - \frac{x^2}{2}\right)\Bigg|_0^1 = \frac{3}{2} - y \qquad 0 \le y \le 1$$

Para ver si las dos variables del ejemplo 8 son estadísticamente independientes debemos determinar si $f(x, y) = f(x)f(y)$. Como $(2 - x - y) \neq (\frac{3}{2} - x)(\frac{3}{2} - y)$, podemos decir que las dos variables no son estadísticamente independientes.

## A.5 Características de las distribuciones de probabilidad

Una distribución de probabilidades a menudo se resume en términos de algunas de sus características, conocidas como **momentos** de la distribución. Dos de los momentos más comunes son la **media**, o **valor esperado,** y la **varianza**.

### Valor esperado

El valor esperado de una va discreta *X,* denotado por $E(X)$, se define de la siguiente manera:

$$E(X) = \sum_x xf(x)$$

donde $\sum_x$ significa la suma sobre todos los valores de *X* y $f(x)$ es la FDP (discreta) de *X*.

**EJEMPLO 10** Considere la distribución de probabilidades de la suma de dos números en el lanzamiento de dos dados analizada en el ejemplo 2 (figura A.1). Al multiplicar los diversos valores de *X,* dados allí por sus correspondientes probabilidades, y sumar sobre todas las observaciones, obtenemos:

$$E(X) = 2\left(\tfrac{1}{36}\right) + 3\left(\tfrac{2}{26}\right) + 4\left(\tfrac{3}{36}\right) + \cdots + 12\left(\tfrac{1}{36}\right)$$
$$= 7$$

que es el valor promedio de la suma de los números observada en un lanzamiento de dos dados.

**EJEMPLO 11** Estime $E(X)$ y $E(Y)$ para la información del ejemplo 4. Ya vimos que

| $x$ | −2 | 0 | 2 | 3 |
|---|---|---|---|---|
| $f(x)$ | 0.27 | 0.12 | 0.26 | 0.35 |

Por consiguiente,

$$E(X) = \sum_x xf(x)$$
$$= (-2)(0.27) + (0)(0.12) + (2)(0.26) + (3)(0.35)$$
$$= 1.03$$

**EJEMPLO 11** (*continuación*)

En forma similar,

| $y$ | 3 | 6 |
|---|---|---|
| $f(y)$ | 0.51 | 0.49 |

$$E(Y) = \sum_y y f(y)$$
$$= (3)(0.51) + (6)(0.49)$$
$$= 4.47$$

El valor esperado de una va continua se define como

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

La única diferencia entre este caso y el valor esperado de una va discreta es que el símbolo de sumatoria se reemplaza por el símbolo de integral.

**EJEMPLO 12**

Determinemos el valor esperado de la FDP continua del ejemplo 3:

$$E(X) = \int_0^3 x \left(\frac{x^2}{9}\right) dx$$
$$= \frac{1}{9}\left[\left(\frac{x^4}{4}\right)\right]_0^3$$
$$= \frac{9}{4}$$
$$= 2.25$$

## Propiedades de los valores esperados

1. El valor esperado de una constante es la constante misma. Así, si $b$ es una constante, $E(b) = b$.
2. Si $a$ y $b$ son constantes,

$$E(aX + b) = aE(X) + b$$

Esto se puede generalizar. Si $X_1, X_2, \ldots, X_N$ son $N$ variables aleatorias y $a_1, a_2, \ldots a_N$ y $b$ son constantes,

$$E(a_1X_1 + a_2X_2 + \cdots + a_NX_N + b) = a_1E(X_1) + a_2E(X_2) + \cdots + a_NE(X_N) + b$$

3. Si $X$ y $Y$ son variables aleatorias *independientes*,

$$E(XY) = E(X)E(Y)$$

Es decir, la esperanza del producto $XY$ es el producto de las esperanzas individuales de $X$ y $Y$.

Sin embargo, observe que

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

aunque $X$ y $Y$ sean independientes.

4. Si $X$ es una variable aleatoria con FDP $f(x)$ y si $g(X)$ es cualquier función de $X$, entonces

$$E[g(X)] = \sum_x g(X)f(x) \qquad \text{si } X \text{ es discreta}$$
$$= \int_{-\infty}^{\infty} g(X)f(x)\,dx \qquad \text{si } X \text{ es continua}$$

Por tanto, si $g(X) = X^2$,

$$E(X^2) = \sum_x x^2 f(X) \qquad \text{si } X \text{ es discreta}$$
$$= \int_{-\infty}^{\infty} x^2 f(X)\,dx \qquad \text{si } X \text{ es continua}$$

**EJEMPLO 13** Considere la siguiente FDP:

| $x$ | $-2$ | $1$ | $2$ |
|---|---|---|---|
| $f(x)$ | $\frac{5}{8}$ | $\frac{1}{8}$ | $\frac{2}{8}$ |

Entonces,

$$E(X) = -2\left(\tfrac{5}{8}\right) + 1\left(\tfrac{1}{8}\right) + 2\left(\tfrac{2}{8}\right)$$
$$= -\tfrac{5}{8}$$

y

$$E(X^2) = 4\left(\tfrac{5}{8}\right) + 1\left(\tfrac{1}{8}\right) + 4\left(\tfrac{2}{8}\right)$$
$$= \tfrac{29}{8}$$

## Varianza

Sea $X$ una variable aleatoria y sea $E(X) = \mu$. La distribución o dispersión de los valores de $X$ alrededor del valor esperado se mide por la varianza, la cual se define como

$$\text{var}(X) = \sigma_X^2 = E(X - \mu)^2$$

La raíz cuadrada positiva de $\sigma_X^2$, $\sigma_X$, se define como **desviación estándar** de $X$. La varianza o la desviación estándar da una indicación de qué tan cercanos o dispersos están los valores individuales de $X$ respecto del valor de su media.

La varianza definida anteriormente se calcula de la siguiente forma:

$$\text{var}(X) = \sum_x (X - \mu)^2 f(x) \qquad \text{si } X \text{ es una va discreta}$$
$$= \int_{-\infty}^{\infty} (X - \mu)^2 f(x)\,dx \qquad \text{si } X \text{ es una va continua}$$

Por conveniencia de cálculo, la fórmula de la varianza anterior se expresa también como

$$\text{var}(X) = \sigma_x^2 = E(X - \mu)^2$$
$$= E(X^2) - \mu^2$$
$$= E(X^2) - [E(X)]^2$$

Al aplicar esta fórmula se ve que la varianza de la variable aleatoria dada en el ejemplo 13 es $\frac{29}{8} - (-\frac{5}{8})^2 = \frac{207}{64} = 3.23$.

**EJEMPLO 14** Determinemos la varianza de la variable aleatoria del ejemplo 3:

$$\text{var}(X) = E(X^2) - [E(X)]^2$$

Ahora,

$$\begin{aligned} E(X^2) &= \int_0^3 x^2\left(\frac{x^2}{9}\right)dx \\ &= \int_0^3 \frac{x^4}{9}dx \\ &= \frac{1}{9}\left[\frac{x^5}{5}\right]_0^3 \\ &= 243/45 \\ &= 27/5 \end{aligned}$$

Como $E(X) = \frac{9}{4}$ (ejemplo 12), tenemos al final

$$\begin{aligned} \text{var}(X) &= 243/45 - \left(\tfrac{9}{4}\right)^2 \\ &= 243/720 = 0.34 \end{aligned}$$

## Propiedades de la varianza

1. $E(X-\mu)^2 = E(X^2) - \mu^2$, como ya mencionamos.
2. La varianza de una constante es cero.
3. Si $a$ y $b$ son constantes,

$$\text{var}(aX + b) = a^2 \,\text{var}(X)$$

4. Si $X$ y $Y$ son variables aleatorias *independientes*,

$$\begin{aligned} \text{var}(X+Y) &= \text{var}(X) + \text{var}(Y) \\ \text{var}(X-Y) &= \text{var}(X) + \text{var}(Y) \end{aligned}$$

Esto puede generalizarse a más de dos variables.

5. Si $X$ y $Y$ son va *independientes* y $a$ y $b$ son constantes,

$$\text{var}(aX + bY) = a^2\,\text{var}(X) + b^2\,\text{var}(Y)$$

## Covarianza

Sean $X$ y $Y$ dos va con medias $\mu_x$ y $\mu_y$, respectivamente. Entonces, la **covarianza** entre las dos variables se define como

$$\text{cov}(X, Y) = E\{(X-\mu_x)(Y-\mu_y)\} = E(XY) - \mu_x\mu_y$$

Se observa con facilidad que la varianza de una variable es la covarianza de dicha variable con ella misma.

La covarianza se calcula de la siguiente manera:

$$\begin{aligned} \text{cov}(X, Y) &= \sum_y \sum_x (X-\mu_x)(Y-\mu_y) f(x, y) \\ &= \sum_y \sum_x XY f(x, y) - \mu_x\mu_y \end{aligned}$$

si $X$ y $Y$ son variables aleatorias discretas y

$$\begin{aligned}\operatorname{cov}(X, Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_x)(Y-\mu_y)f(x, y)\,dx\,dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf(x, y)\,dx\,dy - \mu_x\,\mu_y\end{aligned}$$

si $X$ y $Y$ son variables aleatorias continuas.

## Propiedades de la covarianza

1. Si $X$ y $Y$ son independientes, su covarianza es cero, pues

$$\begin{aligned}\operatorname{cov}(X, Y) &= E(XY) - \mu_x\mu_y \\ &= \mu_x\mu_y - \mu_x\mu_y \qquad \text{porque } E(XY) = E(X)E(Y) = \mu_x\mu_y \text{ cuando } X \text{ y } Y \text{ son independientes} \\ &= 0\end{aligned}$$

2.

$$\operatorname{cov}(a + bX, c + dY) = bd\operatorname{cov}(X, Y)$$

donde $a$, $b$, $c$ y $d$ son constantes.

**EJEMPLO 15** Determinemos la covarianza entre las variables aleatorias discretas $X$ y $Y$ cuyas FDP conjuntas son iguales a las del ejemplo 4. Del ejemplo 11, ya sabemos que $\mu_x = E(X)$ 1.03 y que $\mu_y = E(Y) = 4.47$.

$$\begin{aligned}E(XY) &= \sum_y\sum_x XYf(x, y) \\ &= (-2)(3)(0.27) + (0)(3)(0.08) + (2)(3)(0.16) + (3)(3)(0) \\ &\quad + (-2)(6)(0) + (0)(6)(0.04) + (2)(6)(0.10) + (3)(6)(0.35) \\ &= 6.84\end{aligned}$$

Por consiguiente,

$$\begin{aligned}\operatorname{cov}(X, Y) &= E(XY) - \mu_x\mu_y \\ &= 6.84 - (1.03)(4.47) \\ &= 2.24\end{aligned}$$

## Coeficiente de correlación

El coeficiente de correlación (poblacional) $\rho$ (rho) se define como

$$\rho = \frac{\operatorname{cov}(X, Y)}{\sqrt{\{\operatorname{var}(X)\operatorname{var}(Y)\}}} = \frac{\operatorname{cov}(X, Y)}{\sigma_x\sigma_y}$$

Así definido, $\rho$ es una medida de la asociación *lineal* entre dos variables y su valor se sitúa entre $-1$ y $+1$, donde $-1$ indica una perfecta asociación negativa y $+1$ indica una perfecta asociación positiva.

De la fórmula anterior se ve que

$$\operatorname{cov}(X, Y) = \rho\sigma_x\sigma_y$$

**EJEMPLO 16**

Estime el coeficiente de correlación para la información del ejemplo 4.

De las FDPs del ejemplo 11 se ve con claridad que $\sigma_x = 2.05$ y $\sigma_y = 1.50$. Demostramos ya que la cov$(X, Y) = 2.24$. Por consiguiente, al aplicar la fórmula anterior, estimamos $\rho$ como $2.24/(2.05)(1.50) = 0.73$.

*Varianzas de variables correlacionadas*

Sean $X$ y $Y$ dos va. Entonces,

$$\begin{aligned}\operatorname{var}(X+Y) &= \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, Y)\\ &= \operatorname{var}(X) + \operatorname{var}(Y) + 2\rho\sigma_x\sigma_y\\ \operatorname{var}(X-Y) &= \operatorname{var}(X) + \operatorname{var}(Y) - 2\operatorname{cov}(X, Y)\\ &= \operatorname{var}(X) + \operatorname{var}(Y) - 2\rho\sigma_x\sigma_y\end{aligned}$$

Sin embargo, si $X$ y $Y$ son independientes, la cov$(X, Y)$ es cero, en cuyo caso la var$(X + Y)$ y la var$(X - Y)$ son ambas iguales a var$(X)$ + var$(Y)$, como ya mencionamos.

Los resultados anteriores pueden generalizarse de la siguiente manera: sea $\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n$; entonces la varianza de la combinación lineal $\sum X_i$ es

$$\begin{aligned}\operatorname{var}\left(\sum_{i=1}^{n} x_i\right) &= \sum_{i=1}^{n} \operatorname{var} X_i + 2\sum_{i<j}\sum \operatorname{cov}(X_i, X_j)\\ &= \sum_{i=1}^{n} \operatorname{var} X_i + 2\sum_{i<j}\sum \rho_{ij}\sigma_i\sigma_j\end{aligned}$$

donde $\rho_{ij}$ es el coeficiente de correlación entre $X_i$ y $X_j$, y donde $\sigma_i$ y $\sigma_j$ son las desviaciones estándar de $X_i$ y $X_j$.

Por tanto,

$$\begin{aligned}\operatorname{var}(X_1 + X_2 + X_3) &= \operatorname{var} X_1 + \operatorname{var} X_2 + \operatorname{var} X_3 + 2\operatorname{cov}(X_1, X_2)\\ &\quad + 2\operatorname{cov}(X_1, X_3) + 2\operatorname{cov}(X_2, X_3)\\ &= \operatorname{var} X_1 + \operatorname{var} X_2 + \operatorname{var} X_3 + 2\rho_{12}\sigma_1\sigma_2\\ &\quad + 2\rho_{13}\sigma_1\sigma_3 + 2\rho_{23}\sigma_2\sigma_3\end{aligned}$$

donde $\sigma_1$, $\sigma_2$ y $\sigma_3$ son las desviaciones estándar de $X_1$, $X_2$ y $X_3$, respectivamente, y donde $\rho_{12}$ es el coeficiente de correlación entre $X_1$ y $X_2$, $\rho_{13}$ entre $X_1$ y $X_3$, y $\rho_{23}$ entre $X_2$ y $X_3$.

## Esperanza condicional y varianza condicional

Sea $f(x, y)$ la FDP conjunta de las variables aleatorias $X$ y $Y$. La esperanza condicional de $X$, dada $Y = y$, se define como

$$\begin{aligned}E(X \mid Y = y) &= \sum_x x f(x \mid Y = y) && \text{si } X \text{ es discreta}\\ &= \int_{-\infty}^{\infty} x f(x \mid Y = y)\, dx && \text{si } X \text{ es continua}\end{aligned}$$

donde $E(X \mid Y = y)$ significa la esperanza condicional de $X$ dada $Y = y$ y donde $f(x \mid Y = y)$ es la FDP condicional de $X$. La esperanza condicional de $Y$, $E(X \mid Y = x)$ se define en forma similar.

*Esperanza condicional*

Observe que $E(X \mid Y)$ es una variable aleatoria porque es una función de la variable condicionante $Y$. Sin embargo, $E(X \mid Y = y)$, donde $y$ es un valor específico de $Y$, es una constante.

*Varianza condicional*

La varianza condicional de $X$ dada $Y = y$ se define como

$$\begin{aligned}
\operatorname{var}(X \mid Y = y) &= E\{[X - E(X \mid Y = y)]^2 \mid Y = y\} \\
&= \sum_x [X - E(X \mid Y = y)]^2 f(x \mid Y = y) \qquad \text{si } X \text{ es discreta} \\
&= \int_{-\infty}^{\infty} [X - E(X \mid Y = y)]^2 f(x \mid Y = y)\, dx \qquad \text{si } X \text{ es continua}
\end{aligned}$$

**EJEMPLO 17**

Calcule $E(Y \mid X = 2)$ y $\operatorname{var}(Y \mid X = 2)$ para los datos del ejemplo 4.

$$\begin{aligned}
E(Y \mid X = 2) &= \sum_y y f(Y = y \mid X = 2) \\
&= 3f(Y = 3 \mid X = 2) + 6f(Y = 6 \mid X = 2) \\
&= 3(0.16/0.26) + 6(0.10/0.26) \\
&= 4.15
\end{aligned}$$

*Nota:* $f(Y = 3 \mid X = 2) = f(Y = 3, X = 2)/f(X = 2) = 0.16/0.26$, y $f(Y = 6 \mid X = 2) = f(Y = 6, X = 2)/f(X = 2) = 0.10)0.26$, por lo que

$$\begin{aligned}
\operatorname{var}(Y \mid X = 2) &= \sum_y [Y - E(Y \mid X = 2)]^2 f(Y \mid X = 2) \\
&= (3 - 4.15)^2(0.16/0.26) + (6 - 4.15)^2(0.10/0.26) \\
&= 2.13
\end{aligned}$$

## Propiedades de la esperanza y la varianza condicionales

1. Si $f(X)$ es una función de $X$, entonces $E(f(X) \mid X) = f(X)$; es decir, la función de $X$ se comporta como una constante en el cálculo de la esperanza condicional sobre $X$. Por tanto, $[E(X^3 \mid X)] = E(X^3)$; esto se debe a que si conocemos $X$, también conocemos $X^3$.
2. Si $f(X)$ y $g(X)$ son funciones de $X$, entonces

$$E[f(X)Y + g(X) \mid X] = f(X)E(Y \mid X) + g(X)$$

   Por ejemplo, $E[XY + cX^2 \mid X] = XE(Y \mid X) + cX^2$, donde $c$ es una constante.
3. Si $X$ y $Y$ son independientes, $E(Y \mid X) = E(Y)$; es decir, si $X$ y $Y$ son variables aleatorias independientes, la esperanza condicional de $Y$, dada $X$, es la misma que la esperanza incondicional de $Y$.

4. **Ley de las esperanzas iteradas.** Resulta interesante observar la siguiente relación entre la esperanza incondicional de una variable aleatoria $Y$, $E(Y)$ y su esperanza condicional basada en otra variable aleatoria $X$, $E(Y \mid X)$:

$$E(Y) = E_X[E(Y \mid X)]$$

   Lo anterior se conoce como la ley de las esperanzas iteradas, que en el presente contexto establece que la esperanza marginal, o incondicional, de $Y$ es igual a la esperanza de su esperanza condicional; el símbolo $E_X$ denota que la esperanza se calcula sobre los valores de $X$. En forma sencilla, esta ley enuncia que si primero obtenemos $E(Y \mid X)$ como una función de $X$ y toma su valor esperado sobre la distribución de los valores $X$, obtenemos $E(Y)$, la esperanza incondicional de $Y$. El lector puede verificar esto con los datos del ejemplo 4.

   Una implicación de la ley de esperanzas iteradas es que si la media condicional de $Y$ dada $X$ (es decir, $E[Y \mid X]$) es cero, la media (no condicional) de $Y$ también es cero. Esto se desprende de inmediato porque en ese caso

$$E[E(Y \mid X)] = E[0] = 0$$

5. Si $X$ y $Y$ son independientes, $\text{var}(Y \mid X) = \text{var}(Y)$.
6. $\text{var}(Y) = E[\text{var}(Y \mid X)] + \text{var}[E(Y \mid X)]$; es decir, la varianza (incondicional) de $Y$ es igual a la esperanza de la varianza condicional de $Y$ más la varianza de la esperanza condicional de $Y$.

## Momentos superiores de las distribuciones de probabilidad

Aunque la media, la varianza y la covarianza son las medidas-resumen más frecuentes de las FDP univariadas y multivariadas, en ocasiones requerimos considerar momentos de orden mayor de las FDP, como los momentos tercero y cuarto. Los momentos tercero y cuarto de una FDP univariada $f(x)$ alrededor del valor de su media ($\mu$) se definen como

$$\text{Tercer momento:} \quad E(X-\mu)^3$$

$$\text{Cuarto momento:} \quad E(X-\mu)^4$$

En general, el momento $r$-ésimo alrededor de la media se define como

$$r\text{-ésimo momento:} \quad E(X-\mu)^r$$

El tercero y cuarto momentos de una distribución sirven a menudo para estudiar la "forma" de una distribución de probabilidades, en particular su **asimetría**, $S$ (es decir, falta de simetría), y su **apuntamiento** o **curtosis**, $K$ (es decir, altura o aplanamiento), como se aprecia en la figura A.3.

Una medida de asimetría se define como

$$S = \frac{E(X-\mu)^3}{\sigma^3} = \frac{\text{tercer momento alrededor de la media}}{\text{desviación estándar elevada al cubo}}$$

Una medida común de curtosis está dada por

$$K = \frac{E(X-\mu)^4}{[E(X-\mu)^2]^2} = \frac{\text{cuarto momento alrededor de la media}}{\text{segundo momento elevado al cuadrado}}$$

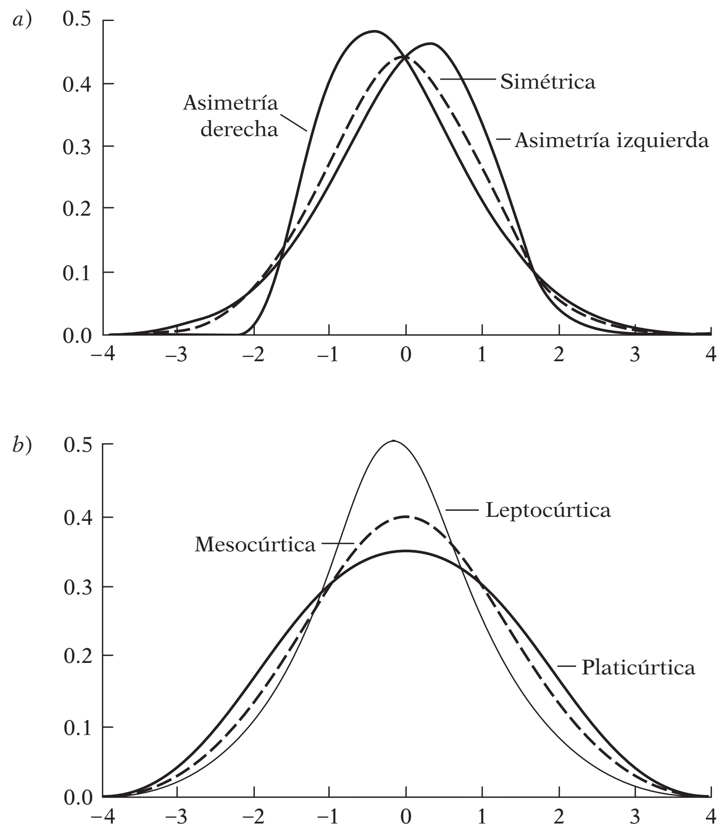



**FIGURA A.3**
*a*) Asimetría;
*b*) curtosis.

Las FDP con valores de *K* menores que 3 se denominan **platicúrticas** (anchas o de colas cortas), y las que tienen valores mayores que 3 se denominan **leptocúrticas** (delgadas o de colas largas). Vea la figura A.3. Una FDP con un valor de curtosis de 3 se conoce como **mesocúrtica**, cuyo ejemplo principal es la distribución normal. (El análisis de la distribución normal está en la sección A.6.)

Más adelante veremos cómo se combinan las medidas de asimetría y curtosis para determinar si una variable aleatoria sigue una distribución normal. Recuerde que el procedimiento de prueba de hipótesis, seguido en las pruebas *t* y *F*, parte del supuesto (por lo menos en muestras pequeñas o finitas) de que la distribución implícita de la variable (o estadístico muestral) es normal. Por consiguiente, es muy importante averiguar si este supuesto se cumple en aplicaciones concretas.

## A.6 Algunas distribuciones de probabilidad teóricas importantes

En el texto se hace uso extenso de las siguientes distribuciones de probabilidad.

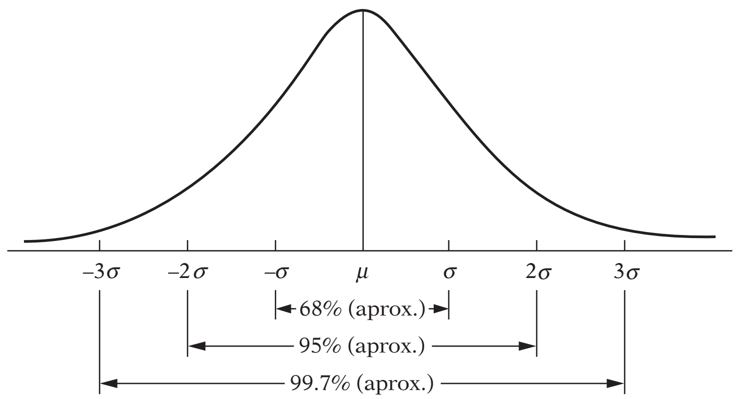
### Distribución normal

La más conocida de todas las distribuciones de probabilidad teóricas es la distribución normal, cuya forma de campana es familiar para quien tenga un mínimo conocimiento estadístico.

Se dice que una variable aleatoria (continua) *X* está normalmente distribuida si su FDP tiene la siguiente forma:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \qquad -\infty < x < \infty$$

**FIGURA A.4**
Áreas bajo la curva normal.




donde $\mu$ y $\sigma^2$, conocidos como *parámetros de la distribución*, son la media y la varianza de la distribución, respectivamente. Las propiedades de esta distribución son las siguientes:

1. Es simétrica alrededor de su valor medio.
2. Aproximadamente 68% del área bajo la curva normal se encuentra entre los valores de $\mu \pm \sigma$, alrededor de 95% del área se encuentra entre $\mu \pm 2\sigma$ y alrededor de 99.7% del área se encuentra entre $\mu \pm 3\sigma$, como se ve en la figura A.4.
3. La distribución normal depende de dos parámetros, $\mu$ y $\sigma^2$. Por tanto, una vez especificados, podemos encontrar la probabilidad de que $X$ esté dentro de cierto intervalo mediante la FDP de la distribución normal. Pero esta labor se aligera considerablemente al consultar la tabla D.1 del **apéndice D**. Para utilizar esta tabla, convertimos la variable dada $X$ normalmente distribuida con media $\mu$ y $\sigma^2$ en una **variable *Z* normal estandarizada** mediante la siguiente transformación:

$$Z = \frac{x - \mu}{\sigma}$$

Una propiedad importante de toda variable estandarizada es que su valor medio es cero y su varianza es la unidad. Así, $Z$ tiene media cero y varianza unitaria. Al sustituir $z$ en la FDP anterior obtenemos

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Z^2\right)$$

que es la FDP de la variable normal estandarizada. Las probabilidades en el **apéndice D**, tabla D.1, se basan en esta variable normal estandarizada.

Por convención, denotamos una variable distribuida normalmente como

$$X \sim N(\mu, \sigma^2)$$

donde $\sim$ significa "distribuido como", $N$ significa distribución normal y las cantidades en los paréntesis son los dos parámetros de la distribución normal, a saber: la media y la varianza. Según esta convención,

$$X \sim N(0, 1)$$

significa que $X$ es una variable normalmente distribuida con media cero y varianza unitaria. En otras palabras, es una variable $Z$ normal estandarizada.

**EJEMPLO 18**

Suponga que $X \sim N(8, 4)$. ¿Cuál es la probabilidad de que $X$ tome un valor entre $X_1 = 4$ y $X_2 = 12$? Para calcular la probabilidad requerida, obtenemos los valores $Z$ como

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{4 - 8}{2} = -2$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12 - 8}{2} = +2$$

Ahora, de la tabla D.1 observamos que $\Pr(0 \le Z \le 2) = 0.4772$. Entonces, por simetría, tenemos $\Pr(-2 \le Z \le 0) = 0.4772$. Por consiguiente, la probabilidad requerida es $0.4772 + 0.4772 = 0.9544$. (Figura A.4.)

**EJEMPLO 19**

¿Cuál es la probabilidad de que, en el ejemplo anterior, $X$ exceda 12?

Esta probabilidad es la misma de que $Z$ exceda 2. De la tabla D.1, es obvio que esta probabilidad es $(0.5 - 0.4772)$ o 0.0228.

4. Sea $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$ y suponga que son independientes. Considere ahora la combinación lineal

$$Y = aX_1 + bX_2$$

donde $a$ y $b$ son constantes. Entonces se demuestra que

$$Y \sim N\left[(a\mu_1 + b\mu_2), \left(a^2\sigma_1^2 + b^2\sigma_2^2\right)\right]$$

Este resultado, que establece que *una combinación lineal de variables normalmente distribuidas también tiene distribución normal*, se generaliza fácilmente a una combinación lineal de más de dos variables normalmente distribuidas.

5. **Teorema central del límite**. Sean $X_1, X_2, \ldots, X_n$ $n$ variables aleatorias independientes, las cuales tienen la misma FDP con media $= \mu$ y varianza $= \sigma^2$. Sea $\bar{X} = \sum X_i/n$ (es decir, la media muestral). Entonces, a medida que $n$ aumenta indefinidamente (es decir, $n \to \infty$),

$$\underset{n\to\infty}{\bar{X}} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Es decir, $\bar{X}$ se acerca a la distribución normal con media $\mu$ y varianza $\sigma^2/n$. Observe que este resultado se cumple sin importar la forma de la FDP. Como resultado, se deduce que

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - u)}{\sigma} \sim N(0, 1)$$

Es decir, $Z$ es una variable normal estandarizada.

6. Los momentos tercero y cuarto de la distribución normal alrededor del valor de la media son los siguientes:

Tercer momento: $E(X - \mu)^3 = 0$

Cuarto momento: $E(X - \mu)^4 = 3\sigma^4$

*Nota:* Todos los momentos elevados a potencias impares alrededor del valor de la media de una variable normalmente distribuida son cero.

7. Como resultado, y según las medidas de asimetría y curtosis ya analizadas, para una FDP normal tenemos una asimetría = 0 y curtosis = 3; es decir, una distribución normal es simétrica y

mesocúrtica. Por consiguiente, una prueba simple de normalidad es determinar si los valores calculados de asimetría y curtosis se apartan de las normas de 0 y 3. Ésta es, en realidad, la lógica de la **prueba de normalidad de Jarque-Bera (JB)** estudiada en el texto:

$$\text{JB} = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \tag{5.12.1}$$

donde $S$ significa asimetría y $K$ curtosis. Según la hipótesis nula de normalidad, JB está distribuida como un estadístico **ji cuadrada** con 2 gl.

8. La media y la varianza de una variable aleatoria normalmente distribuida son independientes, en el sentido de que la primera no es una función de la segunda.
9. Si $X$ y $Y$ están distribuidas normalmente de manera conjunta, son independientes si y sólo si la covarianza entre ellas [es decir, $\operatorname{cov}(X, Y)$] es cero. (Consulte el ejercicio 4.1.)

## Distribución $\chi^2$ (ji cuadrada)

Sean $Z_1, Z_2, \ldots, Z_k$ variables normales estandarizadas *independientes* (es decir, variables normales con media cero y varianza unitaria). Así, se dice que la cantidad

$$Z = \sum_{i=1}^{k} Z_i^2$$

sigue la distribución $\chi^2$ con $k$ grados de libertad (gl), donde el término gl significa el número de cantidades independientes en la suma anterior. Una variable distribuida como ji cuadrada se denota por $\chi_k^2$, donde el subíndice $k$ indica los gl. Geométricamente, la distribución ji cuadrada aparece en la figura A.5.

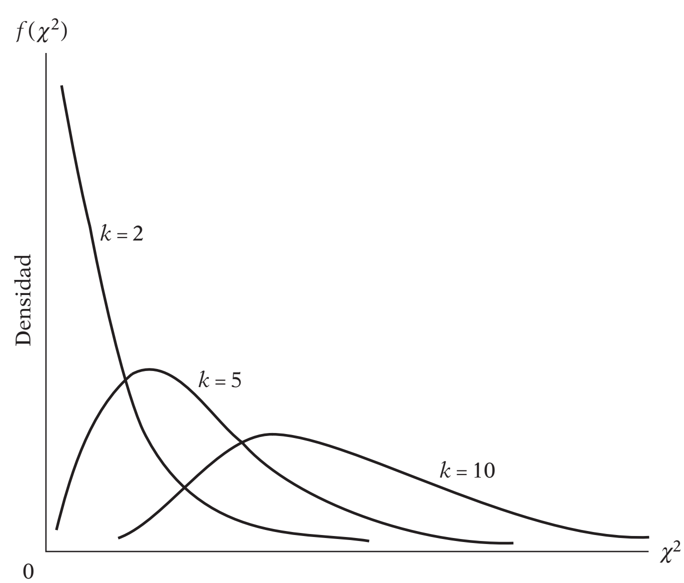
Las propiedades de la distribución $\chi^2$ son las siguientes:

1. Como indica la figura A.5, la distribución $\chi^2$ es una distribución asimétrica; el grado de la asimetría depende de los gl. Cuando los gl son comparativamente pocos, la distribución está muy sesgada hacia la derecha; pero a medida que aumenta el número de gl, la distribución es cada vez más simétrica. De hecho, para gl por encima de 100, la variable

$$\sqrt{2\chi^2} - \sqrt{(2k-1)}$$

puede tratarse como una variable normal estandarizada, donde $k$ son los gl.

**FIGURA A.5**
Función de densidad de la variable $\chi^2$.



2. La media de la distribución ji cuadrada es $k$ y su varianza es $2k$, donde $k$ son los gl.
3. Si $Z_1$ y $Z_2$ son dos variables ji cuadrada independientes con $k_1$ y $k_2$ gl, la suma $Z_1 + Z_2$ es también una variable ji cuadrada con gl $= k_1 + k_2$.

**EJEMPLO 20**

¿Cuál es la probabilidad de obtener un valor $\chi^2$ de 40 o superior, si los gl son 20?

Como muestra la tabla D.4, la probabilidad de obtener un valor $\chi^2$ de 39.9968 o mayor (20 gl) es 0.005. Por consiguiente, la probabilidad de obtener un valor $\chi^2$ de 40 es menor que 0.005, probabilidad un tanto baja.

## Distribución *t* de Student

Si $Z_1$ es una variable normal estandarizada [es decir, $Z_1 \sim N(0, 1)$] y otra variable $Z_2$ sigue la distribución ji cuadrada con $k$ gl y está distribuida independientemente de $Z_1$, entonces la variable definida como

$$t = \frac{Z_1}{\sqrt{(Z_2/k)}} = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}}$$

sigue la distribución $t$ de Student con $k$ gl. Una variable distribuida $t$ se designa con frecuencia como $t_k$, donde el subíndice $k$ denota los gl. Geométricamente, la distribución $t$ se muestra en la figura A.6.

Las propiedades de la distribución $t$ de Student son las siguientes:

1. Como indica la figura A.6, la distribución $t$, lo mismo que la distribución normal, es simétrica, pero es más plana que la normal. Sin embargo, a medida que aumentan los gl, la distribución $t$ se aproxima a la distribución normal.
2. La media de la distribución $t$ es cero y su varianza es $k/(k-2)$.

La distribución $t$ está tabulada en la tabla D.2.

**EJEMPLO 21**

Si gl $= 13$, ¿cuál es la probabilidad de obtener un valor $t$ (*a*) de 3 o más, (*b*) de alrededor de $-3$ o más pequeño y (*c*) de $|t|$ de alrededor de 3 o superior, donde $|t|$ significa el valor absoluto (es decir, sin tomar en cuenta el signo) de $t$?

De la tabla D.2, las respuestas son (*a*) alrededor de 0.005, (*b*) alrededor de 0.005 debido a la simetría de la distribución $t$ y (*c*) alrededor de 0.01 = 2(0.005).

**FIGURA A.6**
Distribución $t$ de Student para grados de libertad seleccionados.

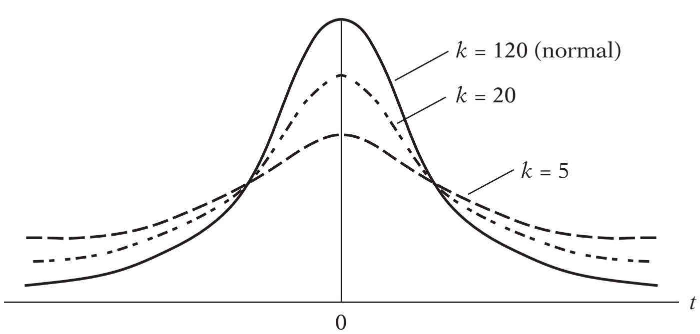
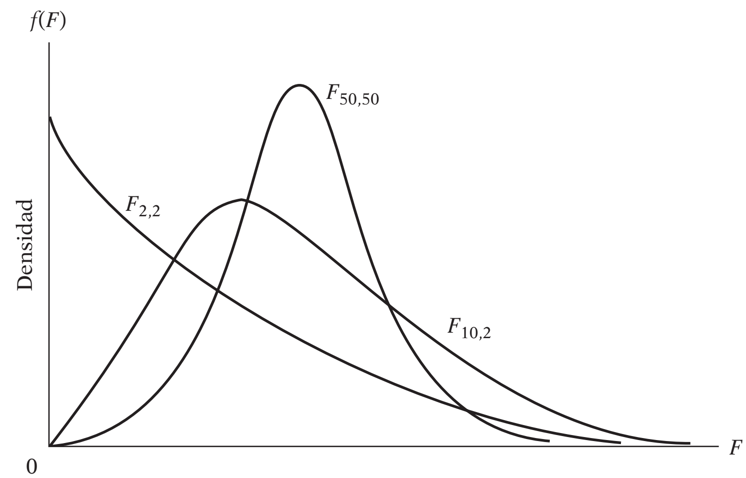


**FIGURA A.7**
Distribución F para diversos grados de libertad.




## Distribución *F*

Si $Z_1$ y $Z_2$ son variables ji cuadrada distribuidas en forma independiente con $k_1$ y $k_2$ gl, respectivamente, la variable

$$F = \frac{Z_1/k_1}{Z_2/k_2}$$

sigue la distribución *F* (de Fisher) con $k_1$ y $k_2$ gl. Una variable que sigue una distribución *F* se denota por $F_{k_1,k_2}$, donde los subíndices indican los gl asociados con las dos variables *Z*, y donde $k_1$ son los *gl del numerador* y $k_2$ los *gl del denominador*. En la figura A.7 se muestra geométricamente la distribución *F*.

La distribución *F* tiene las siguientes propiedades:

1. Al igual que la distribución ji cuadrada, la distribución *F* está sesgada hacia la derecha. Pero puede demostrarse que, a medida que aumentan $k_1$ y $k_2$, la distribución *F* se acerca a la distribución normal.
2. El valor de la media de una variable con distribución *F* es $k_2/(k_2 - 2)$, el cual está definido para $k_2 > 2$ y su varianza es

$$\frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$$

definida para $k_2 > 4$.

3. El cuadrado de una variable aleatoria con distribución *t* con *k* gl sigue una distribución *F* con 1 y *k* gl. Simbólicamente,

$$t_k^2 = F_{1,k}$$

**EJEMPLO 22**

Si $k_1 = 10$ y $k_2 = 8$, ¿cuál es la probabilidad de obtener un valor *F* (*a*) de 3.4 o mayor y (*b*) de 5.8 o mayor?

Como muestra la tabla D.3, estas probabilidades son (*a*) aproximadamente 0.05 y (*b*) aproximadamente 0.01.

4. Si el número de gl del denominador, $k_2$, es relativamente alto, se cumple la siguiente relación entre las distribuciones $F$ y ji cuadrada:

$$k_1 F \sim \chi^2_{k1}$$

Es decir, para un número de gl del denominador relativamente grande, los gl del numerador multiplicados por el valor $F$ equivalen aproximadamente a un valor ji cuadrada con los gl del numerador.

**EJEMPLO 23**

Sean $k_1 = 20$ y $k_2 = 120$. El valor $F$ crítico a 5% para estos gl es 1.48. Por consiguiente, $k_1 F =$ (20)(1.48) = 29.6. De la distribución ji cuadrada para 20 gl, el valor crítico ji cuadrada a 5% es alrededor de 31.41.

A propósito, observe que, como para un número grande de gl las distribuciones $t$, ji cuadrada y $F$ se aproximan a la distribución normal, estas tres se conocen como *distribuciones relacionadas con la distribución normal*.

## Distribución binomial de Bernoulli

Se dice que una variable aleatoria $X$ sigue una distribución de Bernoulli (en honor del matemático suizo) si su función de densidad de probabilidad (FDP), o de masa, es:

$$P(X = 0) = 1 - p$$
$$P(X = 1) = p$$

donde $p$, $0 \leq p \leq 1$, es la probabilidad de que algún suceso sea un "éxito", como la probabilidad de que caiga cara en un lanzamiento de moneda. Para tal variable,

$$E(X) = [1 \times p(X = 1) + 0 \times p(X = 0)] = p$$
$$\text{var}\,(X) = pq$$

donde $q = (1 - p)$, es decir, la probabilidad de un "fracaso".

## Distribución binomial

Esta distribución es la generalización de la distribución de Bernoulli. Sea $n$ el número de intentos independientes, cuyos resultados son un "éxito" con una probabilidad $p$ y un "fracaso" con una probabilidad $q = (1 - p)$. Si $X$ representa el número de éxitos en $n$ intentos, se dice que $X$ sigue una distribución binomial cuya FDP es:

$$f(X) = \binom{n}{x} p^x (1 - p)^{n-x}$$

donde $x$ representa el número de éxitos en $n$ intentos y donde

$$\binom{n}{x} = \frac{n!}{x!(n - x)!}$$

donde $n!$ se lee como $n$ factorial, lo cual significa $n(n - 1)(n - 2)\ldots 1$.

La distribución binomial es de dos parámetros, $n$ y $p$. Para dicha distribución,

$$E(X) = np$$
$$\text{var}\,(X) = np(1 - p) = npq$$

Por ejemplo, si usted lanza una moneda 100 veces y desea saber la probabilidad de obtener 60 caras, debe colocar $p = 0.5$, $n = 100$ y $x = 60$ en la fórmula anterior. Existen rutinas de cálculo para evaluar tales probabilidades.

Puede observar que la distribución binomial es una generalización de la distribución de Bernoulli.

### Distribución de Poisson

Se dice que una variable aleatoria $X$ tiene distribución de Poisson si su FDP es:

$$f(X) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{para } x = 0, 1, 2, \ldots, \lambda > 0$$

La distribución de Poisson depende de un solo parámetro, $\lambda$. Una característica distintiva de la distribución de Poisson es que su varianza es igual a su valor esperado, que es $\lambda$. O sea,

$$E(X) = \text{var}(X) = \lambda$$

Con el modelo de Poisson, como vimos en el capítulo sobre modelos de regresión no lineal, se construyen modelos de fenómenos raros o poco frecuentes, como el número de llamadas recibidas en un lapso de 5 minutos, o el número de multas por rebasar el límite de velocidad recibidas en el transcurso de una hora, o el número de patentes registradas por una compañía a lo largo de un año.

## A.7 Inferencia estadística: estimación

En la sección A.6 consideramos diversas distribuciones de probabilidad teóricas. A menudo conocemos o estamos dispuestos a suponer que una variable aleatoria $X$ sigue una distribución de probabilidades particular, pero no conocemos el valor del (los) parámetro(s) de la distribución. Por ejemplo, si $X$ sigue una distribución normal, quizá deseemos conocer el valor de sus dos parámetros, a saber: la media y la varianza. Para estimar las incógnitas, el procedimiento usual es suponer que tenemos una **muestra aleatoria** de tamaño $n$ de la distribución de probabilidades conocida y con la información muestral estimar los parámetros desconocidos.[5] Esto se conoce como **problema de estimación**. En esta sección consideramos este problema con mayor detalle. El problema de estimación se divide en dos categorías: estimación puntual y estimación por intervalos.

### Estimación puntual

Para establecer las ideas, sea $X$ una variable aleatoria con FDP $f(x; \theta)$, donde $\theta$ es el parámetro de la distribución (para simplificar el análisis, suponemos sólo un parámetro desconocido; el análisis se generaliza fácilmente). Suponga que conocemos la forma funcional; es decir, conocemos la FDP teórica, como la distribución $t$, pero no el valor de $\theta$. Por consiguiente, obtenemos una muestra aleatoria de tamaño $n$ para esta FDP conocida y luego elaboramos una función de valores muestrales tal que

$$\hat{\theta} = f(x_1, x_2, \ldots, x_n)$$

proporciona una estimación del verdadero $\theta$. $\hat{\theta}$ se conoce como un **estadístico** o **estimador**, y un valor numérico particular que tome el estimador, como una **estimación**. Observe que $\hat{\theta}$ puede

[5] Sean $X_1, X_2, \ldots, X_n$ $n$ variables aleatorias con FDP conjunta $f(x_1, x_2, \ldots, x_n)$. Si escribimos

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$$

donde $f(x)$ es la FDP común de cada $X$, se dice que $x_1, x_2, \ldots, x_n$ constituyen una muestra aleatoria de tamaño $n$ de una población con FDP $f(x_n)$.

tratarse como variable aleatoria porque es una función de la información muestral. $\hat{\theta}$ proporciona una regla o fórmula que indica la forma de estimar el verdadero $\theta$. Así, si

$$\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{X}$$

donde $\bar{X}$ es la media muestral, entonces $\bar{X}$ es un estimador del verdadero valor de la media, es decir, $\mu$. Si en un caso específico $\bar{X} = 50$, esto proporciona una estimación de $\mu$. El estimador $\hat{\theta}$ obtenido antes se conoce como estimador puntual porque proporciona sólo una estimación (puntual) de $\theta$.

## Estimación por intervalos

En lugar de obtener sólo una estimación puntual de $\theta$, suponga que obtenemos dos estimaciones de $\theta$ al construir dos estimadores $\hat{\theta}_1(x_1, x_2, \ldots, x_n)$ y $\hat{\theta}_2(x_1, x_2, \ldots, x_n)$, y se dice con alguna confianza (es decir, probabilidad) que el intervalo entre $\hat{\theta}_1$ y $\hat{\theta}_2$ incluye el verdadero $\theta$. Por tanto, en la estimación por intervalos, en contraste con la estimación puntual, se proporciona un intervalo de posibles valores dentro de los cuales se encuentra el verdadero $\theta$.

El concepto clave implícito en la estimación por intervalos es la noción de **muestreo**, o de **distribución de probabilidades, de un estimador**. Por ejemplo, puede demostrarse que si una variable $X$ está normalmente distribuida, entonces la media muestral $\bar{X}$ también está normalmente distribuida con media $= \mu$ (la verdadera media) y varianza $= \sigma^2/n$, donde $n$ es el tamaño de la muestra. En otras palabras, la distribución muestral, o de probabilidad, de un estimador $\bar{X}$ es $\bar{X} \sim N(\mu, \sigma^2/n)$. Como resultado, si construimos el intervalo

$$\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$$

y decimos que hay una probabilidad de aproximadamente 0.95, o 95%, de que intervalos como éste incluyan la verdadera $\mu$, estamos construyendo un estimador por intervalos de $\mu$. Observe que el intervalo anterior es aleatorio, pues se basa en $\bar{X}$, la cual variará de muestra en muestra.

En general, en la estimación por intervalos construimos dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$, ambos funciones de los valores muestrales de $X$, de forma que

$$\Pr(\hat{\theta}_1 \le \theta \le \hat{\theta}_2) = 1 - \alpha \qquad 0 < \alpha < 1$$

Es decir, podemos plantear que la probabilidad es $1 - \alpha$ de que el intervalo de $\hat{\theta}_1$ a $\hat{\theta}_2$ contenga el verdadero $\theta$. Este intervalo se conoce como **intervalo de confianza** de tamaño $1 - \alpha$ para $\theta$, con $1 - \alpha$ como **coeficiente de confianza**. Si $\alpha = 0.05$, entonces $1 - \alpha = 0.95$, lo cual significa que si construimos un intervalo de confianza con un coeficiente de confianza de 0.95, entonces en construcciones repetidas como ésta, resultantes de un muestreo repetido, acertaremos 95 de cada 100 veces si sostenemos que el intervalo contiene el verdadero $\theta$. Cuando el coeficiente de confianza es 0.95, se dice con frecuencia que se tiene un intervalo de confianza a 95%. En general, si el coeficiente de confianza es $1 - \alpha$, se dice que se tiene un intervalo de confianza a $100(1 - \alpha)$%. Observe que $\alpha$ se conoce como **nivel de significancia**, o probabilidad de cometer un error tipo I. Analizamos este tema en la sección A.8.

**EJEMPLO 24**

Suponga que la distribución de las estaturas de los hombres en una población está normalmente distribuida con media $= \mu$ pulgadas y $\sigma = 2.5$ pulgadas (6.35 centímetros). Una muestra de 100 hombres de esta población obtenida al azar tuvo una estatura promedio de 67 pulgadas (170.18 centímetros). Establezca un intervalo de confianza a 95% para la estatura media ($= \mu$) para la población en su conjunto.

Como ya mencionamos, $\bar{X} \sim N(\mu, \sigma^2/n)$ en este caso se convierte en $\bar{X} \sim N(\mu, 2.5^2/100)$. En la tabla D. 1 se ve que

$$\bar{X} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

**EJEMPLO 24** (*continuación*)

cubre 95% del área bajo la curva normal. Por consiguiente, el intervalo anterior proporciona un intervalo de confianza a 95% para $\mu$. Al reemplazar los valores dados de $\bar{X}$, $\sigma$ y $n$ obtenemos el siguiente intervalo de confianza a 95%:

$$66.51 \leq \mu \leq 67.49$$

En mediciones repetidas, como ésta, los intervalos así establecidos incluyen la verdadera $\mu$ con una confianza de 95%. Aquí podemos mencionar un punto técnico: aunque es posible decir que la probabilidad de que el intervalo aleatorio $[\bar{X} \pm 1.96(\sigma/\sqrt{n})]$ incluya $\mu$ es de 95%, *no* podemos afirmar que hay una probabilidad de 95% de que el intervalo particular (66.51, 67.49) incluya $\mu$. Una vez fijado este intervalo, la probabilidad de que incluya $\mu$ *es* de 0 o de 1. Lo que podemos decir es que si construimos 100 intervalos como éste, 95 de los 100 intervalos incluirán la verdadera $\mu$; no podemos garantizar que un intervalo particular necesariamente incluya a $\mu$.

## Métodos de estimación

En términos generales, existen tres métodos de estimación de parámetros: 1) mínimos cuadrados (MC), 2) máxima verosimilitud (MV) y 3) método de los momentos (MOM), y su extensión, el método de los momentos generalizado (MMG). Se ha dedicado mucho tiempo a ilustrar el método de MC. En el capítulo 4 presentamos el método de MV dentro del contexto de la regresión. Pero el método tiene un número mucho mayor de aplicaciones.

La idea central de MV es la **función de verosimilitud**. Para ilustrar lo anterior, suponga que la variable aleatoria $X$ tiene una FDP $f(X, \theta)$ que depende de un solo parámetro $\theta$. Conocemos la FDP (por ejemplo, la de Bernoulli o la binomial), pero no el valor del parámetro. Suponga que obtenemos una muestra aleatoria de $n$ valores $X$. La FDP conjunta para estos $n$ valores es:

$$g(x_1, x_2, \ldots, x_n; \theta)$$

Como es una muestra aleatoria, escribimos la anterior FDP conjunta como el producto de la FDP individual de la siguiente forma:

$$g(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

La FDP conjunta tiene una doble interpretación. Si se conoce $\theta$, la interpretamos como la probabilidad conjunta de observar los valores dados de las muestras. Por otra parte, podemos considerarla una función de $\theta$ para los valores dados de $x_1, x_2, \ldots, x_n$. En esta última interpretación, la FDP se conoce como **función de verosimilitud (FV)** y se expresa como:

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

Observe el papel inverso que desempeña $\theta$ en la función de densidad de probabilidad conjunta y en la función de verosimilitud.

El estimador MV de $\theta$ es el valor de $\theta$ que maximiza la función de verosimilitud (muestra), $L$. Por conveniencia matemática se suele tomar el log de la función de verosimilitud, el cual se conoce como **función logarítmica de verosimilitud (log $L$)**. Según las reglas de cálculo para la maximización, se diferencia la función logarítmica de similitud respecto de la variable desconocida, y la derivada así obtenida se iguala a cero. El valor resultante del estimador se llama **estimador de máxima verosimilitud**. Se puede aplicar la condición de segundo orden de maximización con el objeto de asegurar que el valor obtenido sea en efecto el valor máximo.

En caso de que haya más de un parámetro desconocido, se diferencia la función logarítmica de verosimilitud respecto de cada parámetro desconocido, se igualan los resultados a cero y se resuelven de manera simultánea a fin de obtener los valores de los parámetros desconocidos. Ya hicimos algo similar para el modelo de regresión múltiple (véase el capítulo 4, apéndice 4A.1).

**EJEMPLO 25**

Suponga que la variable aleatoria $X$ sigue la distribución de Poisson y tiene una media igual a $\lambda$. Suponga que $x_1, x_2, \ldots, x_n$ son variables aleatorias independientes con distribución de Poisson y cuya media es $\lambda$. Suponga que queremos calcular el estimador MV de $\lambda$. La función de verosimilitud aquí es:

$$L(x_1, x_2, \ldots, x_n; \lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{x_1!x_2!\cdots x_n!}$$

La anterior es más bien una expresión difícil de manejar, pero si tomamos su logaritmo, se convierte en:

$$\log(x_1, x_2, \ldots, x_n; \lambda) = -n\lambda + \sum x_i \log\lambda - \log c$$

donde $\log c = \prod x_i!$. Al diferenciar la expresión anterior respecto de $\lambda$ obtenemos $(-n + (\sum x_i)/\lambda)$. Si igualamos esta última expresión a cero, tenemos $\lambda_{ml} = (\sum x_i)/n = \bar{X}$, la cual es el estimador MV del parámetro desconocido $\lambda$.

*El método de los momentos*

Ya dimos un vistazo al MOM en el ejercicio 3.4 con el llamado **principio de analogía**, en el cual los momentos muestrales intentan duplicar las propiedades de sus contrapartes poblacionales. El MMG, una generalización del MOM, se populariza cada vez más, pero no en su nivel básico. Por tanto, no se estudia aquí.

Las propiedades estadísticas deseables se encuentran en dos categorías: propiedades de muestra pequeña o muestra finita y propiedades de muestra grande o asintóticas. En estos dos conjuntos de propiedades está implícita la noción de que un estimador tiene una distribución muestral o de probabilidad.

## Propiedades de las muestras pequeñas

*Insesgamiento*

Se dice que un estimador $\hat{\theta}$ es un estimador insesgado de $\theta$ si el valor esperado de $\hat{\theta}$ es igual al verdadero $\theta$; es decir,

$$E(\hat{\theta}) = \theta$$

o

$$E(\hat{\theta}) - \theta = 0$$

Si no se mantiene esta igualdad, decimos que el estimador es sesgado y el sesgo se calcula como

$$\text{sesgo}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Desde luego, si $E(\hat{\theta}) = \theta$ —es decir, $\hat{\theta}$ es un estimador insesgado—, el sesgo es cero.

La situación se ilustra geométricamente en la figura A.8. A propósito, observe que el insesgamiento es una propiedad del muestreo repetido, no de una muestra dada: mantenemos fijo el tamaño de la muestra, obtenemos diversas muestras y cada vez conseguimos una estimación del parámetro desconocido. Se espera que el valor promedio de estas estimaciones sea igual al verdadero valor si el estimador es insesgado.

*Varianza mínima*

Se dice que $\hat{\theta}_1$ es un estimador de varianza mínima de $\theta$ si la varianza de $\hat{\theta}_1$ es menor o igual que la varianza de $\hat{\theta}_2$, que es cualquier otro estimador de $\theta$. La figura A.9 muestra geométricamente

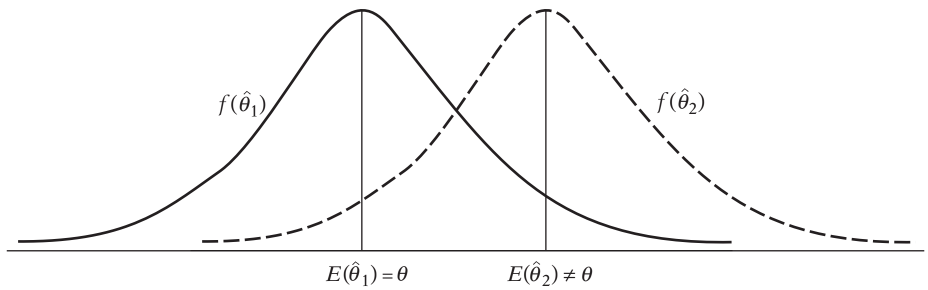



**FIGURA A.8**
Estimadores sesgados e insesgados.

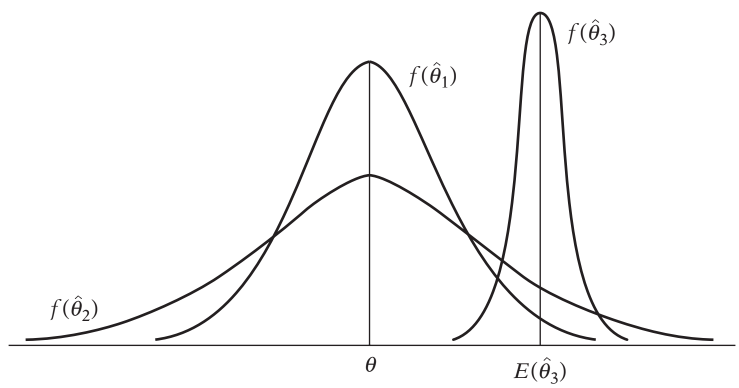



**FIGURA A.9**
Distribución de tres estimadores de $\theta$.

tres estimadores de $\theta$, a saber: $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\theta}_3$, y sus distribuciones de probabilidad. Como se ve, la varianza de $\hat{\theta}_3$ es menor que la de $\hat{\theta}_1$ o la de $\hat{\theta}_2$. Por tanto, si suponemos sólo estos tres estimadores posibles, en este caso $\hat{\theta}_3$ es un estimador de varianza mínima. Pero observe que $\hat{\theta}_3$ es un estimador sesgado (¿por qué?).

### *Mejor estimador insesgado o eficiente*

Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores *insesgados* de $\theta$ y la varianza de $\hat{\theta}_1$ es menor o igual que la varianza de $\hat{\theta}_2$, entonces $\hat{\theta}_1$ es un **estimador insesgado de varianza mínima**, **mejor insesgado** o **eficiente**. Así, en la figura A.9, de los dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$, $\hat{\theta}_1$ es el estimador mejor insesgado o eficiente.

### *Linealidad*

Se dice que un estimador $\hat{\theta}$ es un estimador lineal de $\theta$ si es una función lineal de las observaciones muestrales. Así, la media muestral definida como

$$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

es un estimador lineal porque es una función lineal de los valores de $X$.

### *Mejor estimador lineal insesgado (MELI)*

Si $\hat{\theta}$ es lineal, es insesgado y tiene varianza mínima en la clase de todos los estimadores lineales e insesgados de $\theta$, se denomina **el mejor estimador lineal insesgado**, o **MELI**, para abreviar.

### *Estimador del error cuadrático medio (ECM) mínimo*

El ECM de un estimador $\hat{\theta}$ se define como

$$\text{ECM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

Esto contrasta con la varianza de $\hat{\theta}$, la cual se define como

$$\text{var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

La diferencia entre los dos es que la var($\hat{\theta}$) mide la dispersión de la distribución de $\hat{\theta}$ alrededor de su media o valor esperado, mientras que ECM($\hat{\theta}$) mide la dispersión alrededor del verdadero valor del parámetro. La relación entre los dos es la siguiente:

$$\begin{aligned}
\text{ECM}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\
&= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\
&= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] \\
&= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 \qquad \text{porque el último término es cero}^6 \\
&= \text{var}(\hat{\theta}) + \text{sesgo}(\hat{\theta})^2 \\
&= \text{varianza de } \hat{\theta} \textit{ más} \text{ el sesgo al cuadrado}
\end{aligned}$$

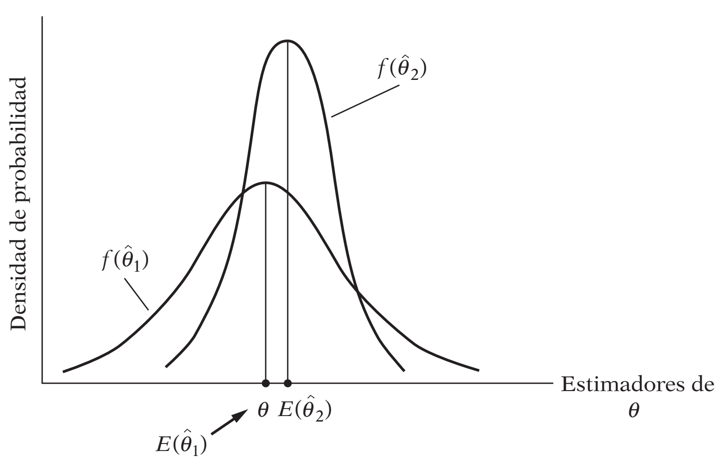
Desde luego, si el sesgo es cero, ECM($\hat{\theta}$) = var($\hat{\theta}$).

El criterio del ECM mínimo consiste en seleccionar un estimador cuyo ECM sea el menor en un conjunto de estimadores comparables. Pero observe que aunque se encontrara tal estimador, hay un costo: para obtener varianza mínima quizá sea necesario aceptar algún sesgo. En forma geométrica, la situación es como se indica en la figura A.10. En ella, $\hat{\theta}_2$ está ligeramente sesgado, pero su varianza es menor que la del estimador insesgado $\hat{\theta}_1$. En la práctica, sin embargo, se emplea el criterio del ECM mínimo cuando el criterio de mejor insesgado es incapaz de producir estimadores con varianzas más pequeñas.

## Propiedades de las muestras grandes

Con frecuencia sucede que un estimador no satisface una o más de las propiedades estadísticas deseables en muestras pequeñas. Pero, a medida que el tamaño de la muestra aumenta indefinidamente, el estimador posee diversas propiedades estadísticas deseables. Estas propiedades se conocen como **propiedades de muestra grande**, o **propiedades asintóticas**.

**FIGURA A.10**
Equilibrio entre sesgo y varianza.




[6] El último término puede escribirse como $2\{[E(\hat{\theta})]^2 - [E(\hat{\theta})]^2 - \theta E(\hat{\theta}) + \theta E(\hat{\theta})\} = 0$. Observe también que $E[E(\hat{\theta}) - \theta]^2 = [E(\hat{\theta}) - \theta]^2$, pues el valor esperado de una constante es tan sólo la constante misma.

*Insesgamiento asintótico*

Se dice que un estimador $\hat{\theta}_1$ es un estimador asintóticamente insesgado de $\theta$ si

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

donde $\hat{\theta}_n$ significa que el estimador está basado en un tamaño de muestra $n$, y donde lím quiere decir límite y $n \to \infty$ significan que $n$ aumenta de manera indefinida. En palabras, $\hat{\theta}$ es un estimador asintóticamente insesgado de $\theta$ si su valor esperado, o media, se aproxima al verdadero valor a medida que el tamaño de la muestra crece cada vez más. Como ejemplo, considere la siguiente medida de varianza muestral de una variable aleatoria $X$:

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

Puede demostrarse que

$$E(S^2) = \sigma^2\left(1 - \frac{1}{n}\right)$$

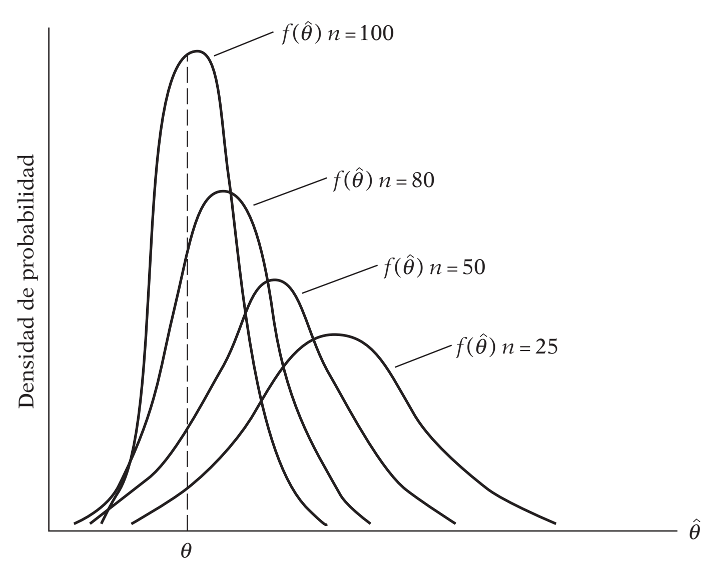
donde $\sigma^2$ es la verdadera varianza. Es obvio que en una muestra pequeña $S^2$ está sesgado, pero, a medida que $n$ aumenta indefinidamente, $E(S^2)$ se aproxima a la verdadera $\sigma^2$; por tanto, éste es asintóticamente insesgado.

*Consistencia*

Se dice que $\hat{\theta}$ es un estimador consistente si se aproxima al verdadero valor de $\theta$ a medida que crece el tamaño de la muestra. La figura A.11 ilustra esta propiedad.

En esta figura tenemos la distribución de $\hat{\theta}$ basada en tamaños muestrales de 25, 50, 80 y 100. Como muestra la figura, $\hat{\theta}$ basada en $n = 25$ está sesgado, pues su distribución muestral no está centrada en el verdadero $\theta$. Pero a medida que $n$ aumenta, la distribución de $\hat{\theta}$ no sólo tiende a centrarse más en $\theta$ (es decir, $\hat{\theta}$ se hace menos sesgada), sino que su varianza también se reduce. Si en el límite (es decir, cuando $n$ aumenta indefinidamente) la distribución de $\hat{\theta}$ se reduce al punto $\theta$, es decir, si la distribución de $\hat{\theta}$ tiene cero dispersión o varianza, se dice que $\hat{\theta}$ es un **estimador consistente** de $\theta$.

**FIGURA A.11**
Distribución de $\hat{\theta}$ a medida que aumenta el tamaño de la muestra.



Más formalmente, se dice que un estimador $\hat{\theta}$ es un estimador consistente de $\theta$ si la probabilidad de que el valor absoluto de la diferencia entre $\hat{\theta}$ y $\theta$ es menor que $\delta$ (una pequeña cantidad positiva arbitraria) y se aproxima a la unidad. Simbólicamente,

$$\lim_{n\to\infty} P\{|\hat{\theta} - \theta| < \delta\} = 1 \qquad \delta > 0$$

donde $P$ significa probabilidad. Esto suele expresarse como

$$\operatorname*{plím}_{n\to\infty} \hat{\theta} = \theta$$

donde plím significa límite de la probabilidad.

Observe que las propiedades de insesgamiento y consistencia son conceptualmente muy diferentes. La propiedad de insesgamiento puede mantenerse para cualquier tamaño de muestra, mientras que la de consistencia es estrictamente una propiedad de muestras grandes.

Una *condición suficiente* para la consistencia es que el sesgo y la varianza tiendan a cero a medida que el tamaño de la muestra aumenta indefinidamente.[7] Otra condición suficiente para la consistencia es que ECM($\hat{\theta}$) tienda a cero a medida que $n$ aumenta de manera indefinida. (Para ECM($\hat{\theta}$), consulte el análisis anterior.)

**EJEMPLO 26**

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media $\mu$ y varianza $\sigma^2$. Demuestre que la media muestral $\bar{X}$ es un estimador consistente de $\mu$.

Por estadística elemental se sabe que $E(\bar{X}) = \mu$ y var($\bar{X}$) $= \sigma^2/n$. Como $E(\bar{X}) = \mu$ sin importar el tamaño de la muestra, ésta es insesgada. Además, a medida que $n$ aumenta indefinidamente, var($\bar{X}$) tiende a cero. Por tanto, $\bar{X}$ es un estimador consistente de $\mu$.

Es importante anotar las siguientes reglas sobre el límite de la probabilidad.

1. *Invarianza (propiedad de Slutsky)*. Si $\hat{\theta}$ es un estimador consistente de $\theta$ y si $h(\hat{\theta})$ es cualquier función continua de $\hat{\theta}$, entonces

$$\operatorname*{plím}_{n\to\infty} h(\hat{\theta}) = h(\theta)$$

Esto significa que si $\hat{\theta}$ es un estimador consistente de $\theta$, entonces $1/\hat{\theta}$ es también un estimador consistente de $1/\theta$ y que $\log(\hat{\theta})$ es también un estimador consistente de $\log(\theta)$. Observe que esta propiedad no se cumple para el operador esperanza $E$; es decir, si $\hat{\theta}$ es un estimador insesgado de $\theta$ [es decir, $E(\hat{\theta}) = \theta$], no es cierto que $1/\hat{\theta}$ sea un estimador insesgado de $1/\theta$; es decir, $E(1/\hat{\theta}) \neq 1/E(\hat{\theta}) \neq 1/\theta$.

2. Si $b$ es una constante, entonces

$$\operatorname*{plím}_{n\to\infty} b = b$$

Es decir, el límite de la probabilidad de una constante es la constante misma.

3. Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores consistentes, entonces

$$\operatorname{plím}(\hat{\theta}_1 + \hat{\theta}_2) = \operatorname{plím}\hat{\theta}_1 + \operatorname{plím}\hat{\theta}_2$$

$$\operatorname{plím}(\hat{\theta}_1\hat{\theta}_2) = \operatorname{plím}\hat{\theta}_1 \operatorname{plím}\hat{\theta}_2$$

$$\operatorname{plím}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) = \frac{\operatorname{plím}\hat{\theta}_1}{\operatorname{plím}\hat{\theta}_2}$$

[7] Más técnicamente, $\text{lím}_{n\to\infty} E(\hat{\theta}_n) = \theta$ y $\text{lím}_{n\to\infty} \text{var}(\hat{\theta}_n) = 0$.

En general, las últimas dos propiedades no se cumplen para el operador de esperanza $E$. Por tanto, $E(\hat{\theta}_1/\hat{\theta}_2) \neq E(\hat{\theta}_1)/(\hat{\theta}_2)$. En forma similar, $E(\hat{\theta}_1\hat{\theta}_2) \neq E(\hat{\theta}_1)E(\hat{\theta}_2)$. Sin embargo, si $\hat{\theta}_1$ y $\hat{\theta}_2$ están distribuidos en forma independiente, $E(\hat{\theta}_1\hat{\theta}_2) = E(\hat{\theta}_1)E(\hat{\theta}_2)$, como ya mencionamos.

*Eficiencia asintótica*

Sea $\hat{\theta}$ un estimador de $\theta$. La varianza de la distribución asintótica de $\hat{\theta}$ se denomina **varianza asintótica** de $\hat{\theta}$. Si $\hat{\theta}$ es consistente y su varianza asintótica es menor que la varianza asintótica de todos los demás estimadores consistentes de $\theta$, $\hat{\theta}$ se llama **asintóticamente eficiente**.

*Normalidad asintótica*

Se dice que un estimador $\hat{\theta}$ está normalmente distribuido asintóticamente si su distribución muestral tiende a aproximarse a la distribución normal a medida que el tamaño de la muestra $n$ aumenta de manera indefinida. Por ejemplo, la teoría estadística muestra que si $X_1, X_2, \ldots, X_n$ son variables independientes normalmente distribuidas con la misma media $\mu$ y la misma varianza $\sigma^2$, la media muestral $\bar{X}$ está también normalmente distribuida con media $\mu$ y varianza $\sigma^2/n$ en muestras pequeñas y en muestras grandes. Pero si las $X_i$ son independientes con media $\mu$ y varianza $\sigma^2$, pero no necesariamente provienen de la distribución normal, entonces la media muestral $\bar{X}$ está normalmente distribuida en forma asintótica con media $\mu$ y varianza $\sigma^2/n$; es decir, a medida que el tamaño de la muestra $n$ aumenta indefinidamente, la media muestral tiende a estar normalmente distribuida con media $\mu$ y varianza $\sigma^2/n$. Ése es, en realidad, el teorema central del límite ya analizado.

## A.8 Inferencia estadística: pruebas de hipótesis

La estimación y las pruebas de hipótesis constituyen ramas gemelas de la inferencia estadística clásica. Una vez examinado el problema de la estimación, consideramos brevemente el problema de pruebas de hipótesis estadísticas.

El problema de pruebas de hipótesis puede plantearse de la siguiente manera: suponga que tenemos una va $X$ con una FDP conocida $f(x; \theta)$, donde $\theta$ es el parámetro de la distribución. Después de obtener una muestra aleatoria de tamaño $n$, obtenemos el estimador puntual $\hat{\theta}$. Como pocas veces conocemos el verdadero $\theta$, planteamos la pregunta: ¿es "compatible" el estimador $\hat{\theta}$ con algún valor de $\theta$ hipotético, por ejemplo, $\theta = \theta^*$, donde $\theta^*$ es un valor numérico específico de $\theta$? En otras palabras, ¿pudo la muestra provenir de FDP $f(x; \theta) = \theta^*$? En el lenguaje de pruebas de hipótesis, $\theta = \theta^*$ se denomina **hipótesis nula** (sostenida) y suele denotarse por $H_0$. La hipótesis nula se prueba contra una **hipótesis alternativa**, denotada por $H_1$, la cual, por ejemplo, puede plantear que $\theta \neq \theta^*$. (*Nota:* En algunos libros de texto, $H_0$ y $H_1$ se designan por $H_1$ y $H_2$, respectivamente.)

La hipótesis nula y la hipótesis alternativa pueden ser **simples** o **compuestas**. Una hipótesis se denomina *simple* si especifica el (los) valor(es) del (los) parámetro(s) de la distribución; de otra forma, se denomina hipótesis *compuesta*. Así, si $X \sim N(\mu, \sigma^2)$ y declaramos que

$$H_0\colon \mu = 15 \qquad \text{y} \qquad \sigma = 2$$

es una hipótesis simple, mientras que

$$H_0\colon \mu = 15 \qquad \text{y} \qquad \sigma > 2$$

es una hipótesis compuesta, porque aquí el valor de $\sigma$ no está especificado.

Para probar la hipótesis nula (es decir, para probar su validez), utilizamos la información muestral con el fin de obtener lo que se conoce como estadístico de prueba. Con mucha frecuencia, este **estadístico de prueba** resulta ser el estimador puntual del parámetro desconocido.

Entonces, tratamos de averiguar la *distribución muestral o probabilística* del estadístico de prueba y utilizar el método de **intervalos de confianza** o de **pruebas de significancia** para probar la hipótesis nula. Ilustraremos este procedimiento más adelante.

Para fijar las ideas, considere de nuevo el ejemplo 23, relacionado con la estatura ($X$) de los hombres en una población. Se nos informó que

$$X_i \sim N(\mu, \sigma^2) = N(\mu, 2.5^2)$$
$$\bar{X} = 67 \qquad n = 100$$

Supongamos que

$$H_0: \mu = \mu^* = 69$$
$$H_1: \mu \neq 69$$

La pregunta es: ¿pudo la muestra con $\bar{X} = 67$, el estadístico de prueba, provenir de la población con el valor de la media de 69? Por intuición, no podemos rechazar la hipótesis nula si $\bar{X}$ está "lo bastante cerca" de $\mu^*$; de lo contrario, podemos rechazarla en favor de una hipótesis alternativa. Pero ¿cómo decidir que $\bar{X}$ está "lo bastante cerca" de $\mu^*$? Podemos adoptar dos métodos: 1) intervalos de confianza y 2) pruebas de significancia, ambos conducentes a conclusiones idénticas en cualquier aplicación específica.

## Método del intervalo de confianza

Como $X_i \sim N(\mu, \sigma^2)$, sabemos que el estadístico de prueba $\bar{X}$ está distribuido como

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Como conocemos la distribución de probabilidades de $\bar{X}$, ¿por qué no establecer, por ejemplo, un intervalo de confianza de $100(1-\alpha)$ para $\mu$ basada en $\bar{X}$ y ver si este intervalo incluye $\mu = \mu^*$? Si es así, no rechazamos la hipótesis nula; si no lo es, la rechazamos. Así, si $\alpha = 0.05$, tendremos un intervalo de confianza a 95%, y si este intervalo de confianza incluye $\mu^*$, no rechazamos la hipótesis nula, pues es probable que 95 de 100 intervalos así construidos incluyan a $\mu^*$.

El procedimiento es el siguiente: como $\bar{X} \sim N(\mu, \sigma^2/n)$, se deduce que

$$Z_i = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

es decir, una variable normal estándar. Entonces, de la tabla de distribución normal, sabemos que

$$\Pr(-1.96 \leq Z_i \leq 1.96) = 0.95$$

Es decir,

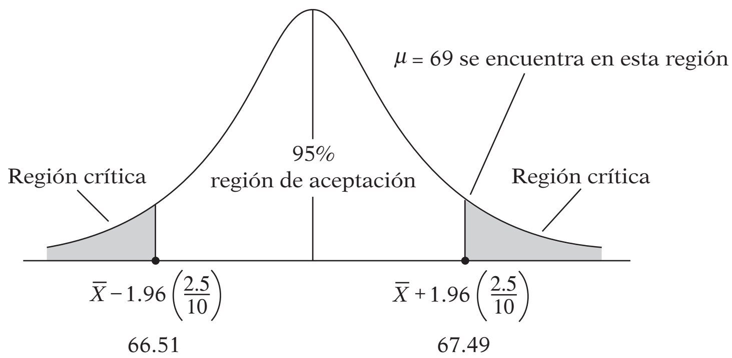
$$\Pr\left(-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

lo cual, al reordenar términos, da

$$\Pr\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

Éste es un intervalo de confianza a 95% para $\mu$. Una vez construido este intervalo, la prueba de la hipótesis nula es simple. Todo lo que debemos hacer es ver si $\mu = \mu^*$ se encuentra en este intervalo. Si se encuentra, podemos aceptar la hipótesis nula; si no se encuentra, la podemos rechazar.

**FIGURA A.12**
Intervalo de confianza a 95% para $\mu$.




De regreso con el ejemplo 24, establecimos ya un intervalo de confianza a 95% para $\mu$, que es

$$66.51 \leq \mu \leq 67.49$$

Como es obvio, este intervalo no incluye $\mu = 69$. Por consiguiente, rechazamos la hipótesis nula de que el verdadero $\mu$ es 69 con un coeficiente de confianza de 95%. La situación se ilustra geométricamente en la figura A.12.

En el lenguaje de pruebas de hipótesis, el intervalo de confianza que construimos se denomina **región de aceptación**, y el(las) área(s) por fuera de la región de aceptación, **región(es) crítica(s)**, o **región(es) de rechazo**, de la hipótesis nula. Los límites inferior y superior de la región de aceptación (que la delimitan con las regiones de rechazo) se denominan **valores críticos**. En este lenguaje de prueba de hipótesis, si el valor hipotético se encuentra dentro de la región de aceptación, podemos aceptar la hipótesis nula; de lo contrario, la podemos rechazar.

Es importante anotar que, en la decisión de rechazar o no $H_0$, es probable que se cometan dos tipos de errores: 1) se puede rechazar $H_0$ cuando es, en realidad, cierta; éste se denomina un **error tipo I** (así, en el ejemplo anterior, $\bar{X} = 67$ pudo provenir de la población con un valor medio de 69), o 2) se puede aceptar $H_0$ cuando, en realidad, es falsa; este error se llama **error tipo II**. Por consiguiente, una prueba de hipótesis no establece el valor de la verdadera $\mu$: tan sólo proporciona un mecanismo para decidir si podemos actuar como si $\mu = \mu^*$.

### *Errores tipo I y tipo II*

Esquemáticamente, tenemos

| | **Estado de naturaleza** | |
|---|---|---|
| **Decisión** | **$H_0$ es verdadera** | **$H_0$ es falsa** |
| Rechazar | Error tipo I | No hay error |
| No rechazar | No hay error | Error tipo II |

Lo ideal sería reducir los errores tipo I y tipo II. Pero, por desgracia, con cualquier tamaño de muestra dado, no es posible reducir ambos errores de manera simultánea. El enfoque clásico de este problema, comprendido en el trabajo de Neyman y Pearson, es suponer la probabilidad de que un error tipo I sea más grave en la práctica que un error tipo II. Por consiguiente, debemos tratar de mantener la probabilidad de cometer un error tipo I en un nivel relativamente bajo, como 0.01 o 0.05, y luego tratar de reducir al máximo la probabilidad de incurrir en un error tipo II.

En la bibliografía, la probabilidad de un error tipo I se representa con $\alpha$ y se denomina **nivel de significancia**, y la probabilidad de un error tipo II se representa con $\beta$. La probabilidad de no cometer un error tipo II se denomina **potencia de la prueba**. Para expresarlo de otra manera: la potencia de una prueba es su capacidad para rechazar una hipótesis nula falsa. El método clásico de pruebas de hipótesis es fijar $\alpha$ en niveles como 0.01 (1%) o 0.05 (5%) y luego tratar de maximizar la potencia de la prueba; es decir, reducir $\beta$.

Es importante que el lector comprenda el concepto de potencia de una prueba, lo cual se explica mejor con un ejemplo.[8]

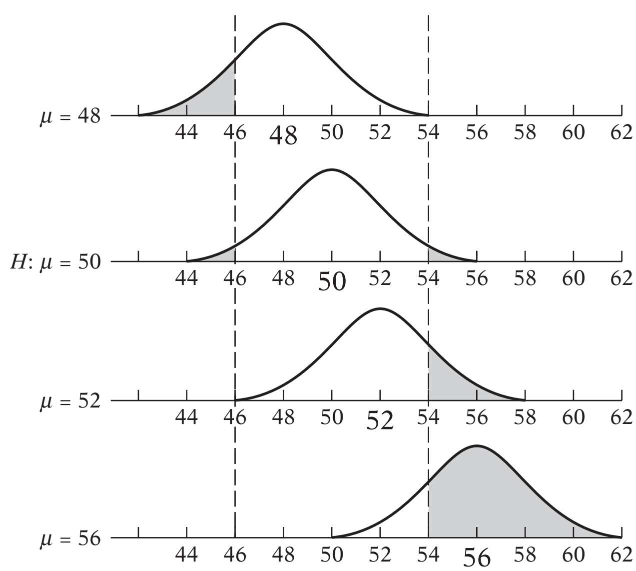
Sea $X \sim N(\mu, 100)$; es decir, $X$ está normalmente distribuida con una media $\mu$ y una varianza 100. Suponga que $\alpha = 0.05$ y que tenemos una muestra de 25 observaciones, la cual da una media muestral de $\bar{X}$. Suponga además que tenemos la hipótesis de que $H_0$: $\mu = 50$. Como $X$ está normalmente distribuida, sabemos que la media muestral está también normalmente distribuida como: $\bar{X} \sim N(\mu,100/25)$. Por tanto, según la hipótesis nula establecida de que $\mu = 50$, el intervalo de confianza de 95% para $\bar{X}$ es $(\mu \pm 1.96(\sqrt{100/25}) = \mu \pm 3.92$, es decir, de 46.08 a 53.92. En consecuencia, la región crítica consiste en todos los valores de $\bar{X}$ menores que 46.08 o mayores que 53.92. Es decir, rechazamos la hipótesis nula de que la verdadera media es 50 si un valor de la media muestral está por debajo de 46.08 o por arriba de 53.92.

Pero ¿cuál es la probabilidad de que $\bar{X}$ esté en la(s) anterior(es) región(es) crítica(s), si la verdadera $\mu$ tiene un valor distinto de 50? Suponga que existen tres hipótesis: $\mu = 48$, $\mu = 52$ y $\mu = 56$. Si cualquiera de ellas es cierta, será la verdadera media de la distribución de $\bar{X}$. El error estándar es invariante para las tres alternativas, pues se sigue suponiendo que $\sigma^2$ toma el valor de 100.

Las áreas sombreadas de la figura A.13 muestran las probabilidades de que $\bar{X}$ se encuentre en la región crítica, si cada hipótesis alternativa es verdadera. Como se puede verificar, dichas pro-

**FIGURA A.13** Distribución de $X$ cuando $N = 25$, $\sigma = 10$ y $\mu = 48$, 50, 52 o 56. Según la hipótesis $H$: $\mu = 50$, la región crítica con $\alpha = 0.05$ es $\bar{X} < 46.1$ y $-X > 53.9$. El área sombreada indica la probabilidad de que $\bar{X}$ se encuentre en la región crítica. Esta probabilidad es:

0.17 si $\mu = 48$ 0.17 si $\mu = 52$
0.05 si $\mu = 50$ 0.85 si $\mu = 56$




[8] El siguiente análisis y las cifras se basan en la obra de Helen M. Walker y Joseph Lev, *Statistical Inference,* Holt, Rinehart y Winston, Nueva York, 1953, pp. 161-162.

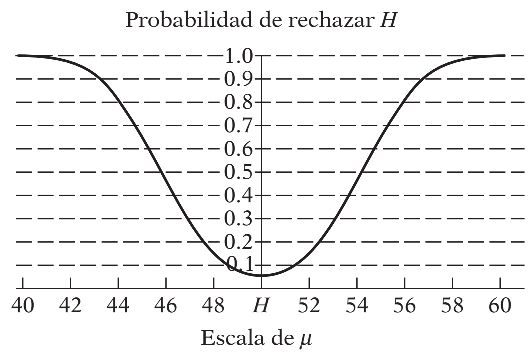



**FIGURA A.14**
Función potencia de la prueba de hipótesis $\mu = 50$ cuando $N = 25$, $\sigma = 10$ y $\alpha = 0.05$.

babilidades son 0.17 (para $\mu = 48$), 0.05 (para $\mu = 50$), 0.17 (para $\mu = 52$) y 0.85 (para $\mu = 56$). Como se observa a partir de esta figura, siempre que el verdadero valor de $\mu$ difiera sustancialmente respecto de la hipótesis en consideración (que aquí es $\mu = 50$), la probabilidad de rechazar la hipótesis es alta, pero cuando el verdadero valor no es muy distinto del valor dado según la hipótesis nula, la probabilidad de rechazo es pequeña. Por intuición, esto debe ser sensato si las hipótesis nula y alternativa están muy próximas entre sí.

Esto resulta más claro si tiene en cuenta la figura A.14, conocida como **gráfica de la función potencia**; asimismo, la curva que ahí se ilustra se conoce como **curva potencia**.

El lector ya se habrá dado cuenta de que el coeficiente de confianza $(1 - \alpha)$ analizado antes es tan sólo uno menos la probabilidad de cometer un error tipo I. Por tanto, un coeficiente de confianza de 95% significa que se está preparado para aceptar, como máximo, una probabilidad de 5% de cometer un error tipo I —no deseamos rechazar la hipótesis verdadera más de 5 veces de cada 100—.

### *Valor p, o nivel exacto de significancia*

En lugar de preseleccionar $\alpha$ en niveles arbitrarios, como 1, 5 o 10%, podemos obtener el **valor *p* (probabilidad)**, o **nivel exacto de significancia** de un estadístico de prueba. El valor $p$ se define como *el nivel de significancia más bajo al cual puede rechazarse una hipótesis nula.*

Suponga que en una aplicación que considera 20 gl obtenemos un valor $t$ de 3.552. Ahora, el valor $p$, o la probabilidad exacta, de obtener un valor $t$ de 3.552 o mayor se ve en la tabla D.2 como 0.001 (a una cola) o 0.002 (a dos colas). Podemos decir que el valor $t$ observado de 3.552 es estadísticamente significativo en el nivel de 0.001 o 0.002, según se emplee una prueba de una o de dos colas.

Diversos paquetes estadísticos imprimen el valor $p$ de los estadísticos de prueba estimados. Por consiguiente, se aconseja al lector dar el valor $p$ siempre que sea posible.

### *Tamaño de la muestra y pruebas de hipótesis*

En datos provenientes de encuestas que comprenden cientos de observaciones, la hipótesis nula se rechaza, al parecer, con mayor frecuencia que en las muestras pequeñas. Aquí vale la pena citar a Angus Deaton:

> A medida que el tamaño de la muestra aumenta, y siempre que se siga un procedimiento de estimación consistente, las estimaciones se aproximarán más a la verdad y estarán menos dispersos alrededor de ésta, de modo que las discrepancias que no pueden detectarse en una muestra de tamaño pequeño provocan rechazo en muestras grandes. Las muestras de tamaño grande son como un poder mayor de resolución en un telescopio; las características que no son visibles a la distancia se vuelven cada vez más nítidas a medida que aumenta la amplificación.[9]

[9] Angus Deaton, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy,* The Johns Hopkins University Press, Baltimore, 2000, p. 130.

Según Leamer y Schwartz, Deaton propone ajustar los valores críticos estándar de las pruebas $F$ y $\chi^2$ como sigue: *Rechace la hipótesis nula cuando el valor F calculado sea superior al logaritmo del tamaño muestral, esto es, ln, y cuando el estadístico $\chi^2$ calculado para la restricción q sea superior a* ***qln,*** *donde l es el logaritmo natural y n es el tamaño de la muestra.* Estos valores críticos se conocen como valores críticos de **Leamer-Schwarz**.

En el ejemplo de Deaton, si $n = 100$, la hipótesis nula se rechazaría sólo si el valor $F$ calculado fuera mayor que 4.6, pero si $n = 10\,000$, la hipótesis nula se rechazaría cuando el valor $F$ calculado fuera superior a 9.2.

## Método de la prueba de significancia

Recuerde que

$$Z_i = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

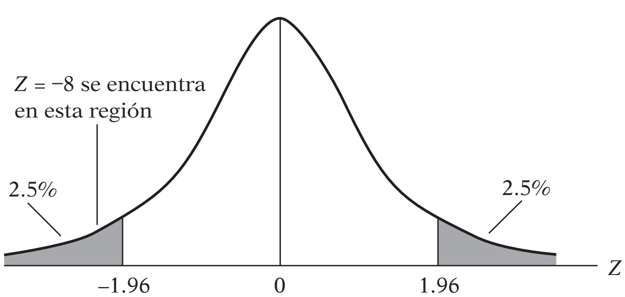
En cualquier aplicación dada, $\bar{X}$ y $n$ se conocen (o se estiman), pero los verdaderos $\mu$ y $\sigma$ no se conocen. Sin embargo, si se especifica $\sigma$ y suponemos (según $H_0$) que $\mu = \mu^*$, un valor numérico específico, entonces $Z_i$ puede calcularse directamente y podemos consultar la tabla de la distribución normal para encontrar la probabilidad de obtener el valor $Z$ calculado. Si esta probabilidad es baja, por ejemplo, menor que 5% o que 1%, podemos rechazar la hipótesis nula: si la hipótesis fuera cierta, la posibilidad de obtener el valor $Z$ particular debería ser muy alta. Ésta es la idea general del método de pruebas de significancia para probar hipótesis. La idea clave es el estadístico de prueba (aquí el estadístico $Z$) y su distribución de probabilidades según el valor supuesto $\mu = \mu^*$. Apropiadamente, en el presente caso, la prueba se conoce como **prueba *Z***, pues utilizamos el valor (normal estandarizado) de $Z$.

Al considerar de nuevo el ejemplo, si $\mu = \mu^* = 69$, el estadístico $Z$ se convierte en

$$\begin{aligned} Z &= \frac{\bar{X} - \mu^*}{\sigma/\sqrt{n}} \\ &= \frac{67 - 69}{2.5/\sqrt{100}} \\ &= -2/0.25 = -8 \end{aligned}$$

En la tabla de la distribución normal D.1 se ve que la probabilidad de obtener ese valor de $Z$ es extremadamente baja. (*Nota:* La probabilidad de que $Z$ exceda 3 o $-3$ es de alrededor de 0.001. Por consiguiente, la probabilidad de que $Z$ exceda 8 es aún menor.) Por tanto, podemos rechazar la hipótesis nula de que $\mu = 69$; con este valor, la probabilidad de obtener una $\bar{X}$ de 67 es extremadamente baja. Así, dudamos que la muestra proviniese de una población con un valor medio de 69. La situación se ilustra en forma diagramática en la figura A.15.

**FIGURA A.15**
Distribución del estadístico $Z$.



En el lenguaje de pruebas de significancia, cuando decimos que una prueba (estadística) es significativa, por lo general nos referimos a que podemos rechazar la hipótesis nula. Y el estadístico de prueba se considera significativo si la probabilidad de obtenerlo es igual o menor que $\alpha$, es decir, la probabilidad de cometer un error tipo I. Así, si $\alpha = 0.05$, sabemos que la probabilidad de obtener un valor $Z$ de $-1.96$ o $1.96$ es de 5% (o 2.5% en cada cola de la distribución normal estandarizada). En el ejemplo ilustrativo, $Z$ era $-8$. Por tanto, la probabilidad de obtener ese valor de $Z$ es muy inferior a 2.5%, muy por debajo de la probabilidad previamente especificada de cometer un error tipo I. Es por esto que el valor calculado de $Z = -8$ es estadísticamente significativo; es decir, rechazamos la hipótesis nula de que la verdadera $\mu^*$ es 69. Por supuesto, llegamos a la misma conclusión con el enfoque de intervalos de confianza para pruebas de hipótesis.

Resumimos ahora los pasos comprendidos en las pruebas de hipótesis estadísticas:

**Paso 1.** Postular la hipótesis nula $H_0$ y la hipótesis alternativa $H_1$ (por ejemplo, $H_0$: $\mu = 69$ y $H_1$: $\mu \neq 69$).

**Paso 2.** Seleccionar el estadístico de prueba (por ejemplo, $\bar{X}$).

**Paso 3.** Determinar la distribución de probabilidades del estadístico de prueba (por ejemplo, $\bar{X} \sim N(\mu,\sigma^2/n)$.

**Paso 4.** Seleccionar el nivel de significancia (es decir, la probabilidad de cometer un error tipo I) $\alpha$.

**Paso 5.** Con la distribución de probabilidades del estadístico de prueba, construir un intervalo de confianza a $100(1 - \alpha)$%. Si el valor del parámetro según la hipótesis nula (por ejemplo, $\mu = \mu^* = 69$) se encuentra en esta región de confianza (la región de aceptación), no se rechaza la hipótesis nula; pero si se encuentra por fuera de este intervalo (es decir, en la región de rechazo), debe rechazar la hipótesis nula. Tenga en mente que al no rechazar o rechazar la hipótesis nula, corre el riesgo de estar equivocado $\alpha$ por ciento de las veces.

## Referencias

Para los detalles del material cubierto en este apéndice, el lector puede consultar las siguientes referencias:

Hoel, Paul G., *Introduction to Mathematical Statistics*, 4a. ed., John Wiley & Sons, Nueva York, 1974. Este libro proporciona una introducción relativamente sencilla a diversos aspectos de la estadística matemática.

Freund, John E. y Ronald E. Walpole, *Mathematical Statistics*, 3a. ed., Prentice Hall, Englewood Cliffs, Nueva Jersey, 1980. Otro libro de texto introductorio en estadística matemática.

Mood, Alexander M., Franklin A. Graybill y Duane C. Boes, *Introduction to the Theory of Statistics*, 3a. ed., McGraw-Hill, Nueva York, 1974. Es una introducción completa a la teoría estadística pero un poco más difícil que los dos libros anteriores.

Newbold, Paul, *Statistics for Business and Economics*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1984. Una introducción no matemática completa a la estadística con muchos problemas resueltos.

Apéndice B

# Nociones básicas de álgebra matricial

En este apéndice se presentan nociones esenciales del álgebra matricial requeridas para entender el apéndice C y parte del material del capítulo 18. El análisis no es riguroso y no se dan pruebas. El lector puede consultar las pruebas y mayores detalles en las referencias.

## B.1 Definiciones

### Matriz

Una matriz es un ordenamiento rectangular de números o de elementos arreglados en renglones y en columnas. Más precisamente, una matriz de **orden**, o de **dimensión**, $M$ por $N$ (escrita como $M \times N$) es un conjunto de $M \times N$ elementos ordenados en $M$ renglones y $N$ columnas. Por tanto, si las letras en negritas denotan matrices, una matriz $\mathbf{A}$ de $(M \times N)$ se expresa como

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{M1} & a_{M2} & a_{M3} & \cdots & a_{MN} \end{bmatrix}$$

donde $a_{ij}$ es el elemento que aparece en el $i$-ésimo renglón y la $j$-ésima columna de $\mathbf{A}$, y donde $[a_{ij}]$ es una expresión abreviada para la matriz $\mathbf{A}$ cuyo elemento característico es $a_{ij}$. El orden o dimensión de una matriz —es decir, el número de filas y columnas— suele escribirse debajo de la matriz como referencia sencilla.

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 2 & 3 & 5 \\ 6 & 1 & 3 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{B}} = \begin{bmatrix} 1 & 5 & 7 \\ -1 & 0 & 4 \\ 8 & 9 & 11 \end{bmatrix}$$

*Escalar*

Un escalar es un número (real). En otro caso, un escalar es una matriz $1 \times 1$.

### Vector columna

Una matriz que consta de $M$ filas y sólo una columna se denomina **vector columna**. Con las letras minúsculas en negritas que denotan vectores, un ejemplo de vector columna es

$$\underset{4\times 1}{\mathbf{x}} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 9 \end{bmatrix}$$

### Vector renglón

Una matriz que consta de sólo un renglón y $N$ columnas se denomina **vector renglón**.

$$\underset{1\times 4}{\mathbf{x}} = [1 \quad 2 \quad 5 \quad -4] \qquad \underset{1\times 5}{\mathbf{y}} = [0 \quad 5 \quad -9 \quad 6 \quad 10]$$

### Trasposición

La traspuesta de una matriz **A** de $M \times N$, denotada por $\mathbf{A}'$ (se lee **A** prima o **A** traspuesta), es una matriz $N \times M$ obtenida mediante el intercambio de renglones y columnas de **A**; es decir, el $i$-ésimo renglón de **A** se convierte en la $i$-ésima columna de $\mathbf{A}'$. Por ejemplo,

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 4 & 5 \\ 3 & 1 \\ 5 & 0 \end{bmatrix} \qquad \underset{2\times 3}{\mathbf{A}'} = \begin{bmatrix} 4 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}$$

Como un vector es un tipo especial de matriz, la traspuesta de un vector renglón es un vector columna, y la de un vector columna un vector renglón. Por tanto,

$$\mathbf{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \qquad \text{y} \qquad \mathbf{x}' = [4 \quad 5 \quad 6]$$

Utilizaremos la convención de indicar los vectores renglón mediante el símbolo de primo.

### Submatriz

Con cualquier matriz **A** de $M \times N$, si se borran todos los renglones y columnas de **A** menos $r$ renglones y $s$ columnas, la matriz resultante de orden $r \times s$ se denomina **submatriz** de **A**. Así, si

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 5 & 7 \\ 8 & 2 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

y borramos el tercer renglón y la tercera columna de **A**, obtenemos

$$\underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 3 & 5 \\ 8 & 2 \end{bmatrix}$$

que es una submatriz de **A** de orden $2 \times 2$.

## B.2 Tipos de matrices

### Matriz cuadrada

Una matriz con el mismo número de renglones y de columnas se denomina **matriz cuadrada**.

$$\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 & 5 & 8 \\ 7 & 3 & 1 \\ 4 & 5 & 0 \end{bmatrix}$$

### Matriz diagonal

Una matriz cuadrada que posee al menos un elemento diferente de cero sobre la diagonal principal (que parte de la esquina superior izquierda hasta la esquina inferior derecha), y con valores restantes de cero, se denomina **matriz diagonal**.

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{B}} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

### Matriz escalar

Una matriz diagonal cuyos elementos diagonales son todos iguales se denomina **matriz escalar**. Un ejemplo es la matriz de varianza-covarianza de las perturbaciones poblacionales del modelo clásico de regresión lineal de la ecuación (C.2.3), a saber:

$$\text{var-cov}\,(\mathbf{u}) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

### Matriz identidad o unitaria

Una matriz diagonal cuyos elementos diagonales son todos 1 se denomina **matriz identidad** o **unitaria** y se denota por **I**. Es una clase especial de matriz escalar.

$$\underset{3\times 3}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \underset{4\times 4}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

### Matriz simétrica

Una matriz cuadrada cuyos elementos por encima de la diagonal son imágenes reflejo de los elementos por debajo de la diagonal principal se denomina **matriz simétrica**. Además, una matriz simétrica es tal que su traspuesta es igual a sí misma; es decir, $\mathbf{A} = \mathbf{A}'$. Es decir, el elemento $a_{ij}$ de **A** es igual al elemento $a_{ji}$ de $\mathbf{A}'$. Un ejemplo es la matriz de varianza-covarianza de la ecuación (C.2.2). Otro ejemplo es la matriz de correlación de (C.5.1).

### Matriz nula

Una matriz cuyos elementos son todos cero se denomina **matriz nula** y se denota por **0**.

### Vector nulo

Un vector renglón o columna cuyos elementos son todos cero se denomina **vector nulo** y se denota también por **0**.

### Matrices iguales

Se dice que dos matrices **A** y **B** son iguales si son del mismo orden y sus elementos correspondientes son iguales; es decir, $a_{ij} = b_{ij}$ para todo $i$ y $j$. Por ejemplo, las matrices

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 4 & 5 \\ 0 & -1 & 2 \\ 5 & 1 & 3 \end{bmatrix} \quad \text{y} \quad \underset{3\times 3}{\mathbf{B}} = \begin{bmatrix} 3 & 4 & 5 \\ 0 & -1 & 2 \\ 5 & 1 & 3 \end{bmatrix}$$

son iguales; es decir, $\mathbf{A} = \mathbf{B}$

## B.3 Operaciones matriciales

### Adición de matrices

Sea $\mathbf{A} = [a_{ij}]$ y $\mathbf{B} = [b_{ij}]$. Si **A** y **B** son del mismo orden, definimos la adición de matrices como

$$\mathbf{A} + \mathbf{B} = \mathbf{C}$$

donde **C** es del mismo orden que **A** y **B** y se obtiene como $c_{ij} = a_{ij} + b_{ij}$ para todo $i$ y $j$; es decir, **C** se obtiene al sumar los elementos correspondientes para la adición. Por ejemplo, si

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & -1 & 3 \\ -2 & 0 & 1 & 5 \end{bmatrix}$$

y **C = A + B,** entonces

$$\mathbf{C} = \begin{bmatrix} 3 & 3 & 3 & 8 \\ 4 & 7 & 9 & 14 \end{bmatrix}$$

## Resta de matrices

La resta de matrices sigue el mismo principio que la adición de matrices, excepto que $\mathbf{C} = \mathbf{A} - \mathbf{B}$; es decir, se restan los elementos de **B** de los elementos correspondientes de **A** para obtener **C**, en tanto **A** y **B** sean del mismo orden.

## Multiplicación por escalar

Para multiplicar una matriz **A** por un escalar λ (un número real), se multiplica cada elemento de la matriz por λ:

$$\lambda\mathbf{A} = [\lambda a_{ij}]$$

Por ejemplo, si $\lambda = 2$ y

$$\mathbf{A} = \begin{bmatrix} -3 & 5 \\ 8 & 7 \end{bmatrix}$$

entonces

$$\lambda\mathbf{A} = \begin{bmatrix} -6 & 10 \\ 16 & 14 \end{bmatrix}$$

## Multiplicación de matrices

Sea **A** $M \times N$ y **B** $N \times P$. Entonces, el producto **AB** (en ese orden) está definido para ser una nueva matriz **C** del orden $M \times P$ tal que

$$c_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj} \qquad \begin{matrix} i = 1, 2, \ldots, M \\ j = 1, 2, \ldots, P \end{matrix}$$

Es decir, el elemento en el $i$-ésimo renglón y la $j$-ésima columna de **C** se obtiene al multiplicar los elementos del $i$-ésimo renglón de **A** por los elementos correspondientes de la $j$-ésima columna de **B** y sumar sobre todos los términos; esto se conoce como regla de la multiplicación del *renglón por columna*. Así, para obtener $c_{11}$, el elemento en el primer renglón y en la primera columna de **C**, se multiplican los elementos en el primer renglón de **A** por los correspondientes en la primera columna de **B** y se suma sobre todos los términos. En forma similar, para obtener $c_{12}$, se multiplican los elementos del primer renglón de **A** por los correspondientes en la segunda columna de **B** y se suma sobre todos los términos, y así sucesivamente.

Observe que, para que exista la multiplicación, las matrices **A** y **B** deben ser conformables respecto de la multiplicación; es decir, el número de columnas en **A** debe ser igual al número de renglones en **B**. Si, por ejemplo,

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 4 & 7 \\ 5 & 6 & 1 \end{bmatrix} \quad \text{y} \quad \underset{3\times 2}{\mathbf{B}} = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 6 & 2 \end{bmatrix}$$

$$\mathbf{AB} = \underset{2\times 2}{\mathbf{C}} = \begin{bmatrix} (3\times 2)+(4\times 3)+(7\times 6) & (3\times 1)+(4\times 5)+(7\times 2) \\ (5\times 2)+(6\times 3)+(1\times 6) & (5\times 1)+(6\times 5)+(1\times 2) \end{bmatrix}$$

$$= \begin{bmatrix} 60 & 37 \\ 34 & 37 \end{bmatrix}$$

Pero si

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 4 & 7 \\ 5 & 6 & 1 \end{bmatrix} \quad \text{y} \quad \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}$$

el producto **AB** no está definido, pues **A** y **B** no son conformables respecto de la multiplicación.

## Propiedades de la multiplicación de matrices

1. La multiplicación de matrices no necesariamente es conmutativa; es decir, en general $\mathbf{AB} \neq \mathbf{BA}$. Por consiguiente, el orden en el cual se multiplican las matrices es muy importante. **AB** significa que **A** es *posmultiplicada* por **B** o **B** es *premultiplicada* por **A**.
2. Aunque existan **AB** y **BA**, las matrices resultantes pueden no ser del mismo orden. Por tanto, si **A** es $M \times N$ y **B** es $N \times M$, **AB** es $M \times M$, mientras que **BA** es $N \times N$, de donde se explica la diferencia de orden.
3. Aunque **A** y **B** sean matrices cuadradas, de manera que **AB** y **BA** estén definidas, las matrices resultantes no necesariamente serán iguales. Por ejemplo, si

$$\mathbf{A} = \begin{bmatrix} 4 & 7 \\ 3 & 2 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} 1 & 5 \\ 6 & 8 \end{bmatrix}$$

entonces

$$\mathbf{AB} = \begin{bmatrix} 46 & 76 \\ 15 & 31 \end{bmatrix} \quad \text{y} \quad \mathbf{BA} = \begin{bmatrix} 19 & 17 \\ 48 & 58 \end{bmatrix}$$

y $\mathbf{AB} \neq \mathbf{BA}$. Un ejemplo de **AB = BA** es cuando tanto **A** como **B** son matrices identidad.

4. Un vector renglón posmultiplicado por un vector columna es un escalar. Por tanto, considere los residuos de mínimos cuadrados ordinarios $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$. Si **u** es un vector columna y **u**′ un vector renglón, tenemos

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= [\hat{u}_1 \quad \hat{u}_2 \quad \hat{u}_3 \quad \cdots \quad \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \vdots \\ \hat{u}_n \end{bmatrix} \\ &= \hat{u}_1^2 + \hat{u}_2^2 + \hat{u}_3^2 + \cdots + \hat{u}_n'^2 \\ &= \sum \hat{u}_i^2 \qquad \text{un escalar [véase la ecuación (C.3.5)]} \end{aligned}$$

5. Un vector columna posmultiplicado por un vector renglón es una matriz. Como ejemplo, considere las perturbaciones poblacionales del modelo clásico de regresión lineal, a saber: $u_1, u_2, \ldots, u_n$. Si **u** es un vector columna y **u**′ un vector renglón, obtenemos

$$\begin{aligned} \mathbf{uu}' &= \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} [u_1 \quad u_2 \quad u_3 \quad \cdots \quad u_n] \\ &= \begin{bmatrix} u_1^2 & u_1u_2 & u_1u_3 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & u_2u_3 & \cdots & u_2u_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_nu_1 & u_nu_2 & u_nu_3 & \cdots & u_n^2 \end{bmatrix} \end{aligned}$$

que es una matriz de orden $n \times n$. Observe que la matriz anterior es simétrica.

6. Una matriz posmultiplicada por un vector columna es un vector columna.
7. Un vector renglón posmultiplicado por una matriz es un vector renglón.
8. La multiplicación de matrices es *asociativa*; es decir, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, donde $\mathbf{A}$ es $M \times N$, $\mathbf{B}$ es $N \times P$ y $\mathbf{C}$ es $P \times K$.
9. La multiplicación de matrices es distributiva respecto de la suma; es decir, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ y $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$.

### Trasposición de matrices

Definimos ya el proceso de trasposición de matrices como el intercambio de renglones y de columnas de una matriz (o de un vector). Ahora presentamos algunas propiedades de la trasposición.

1. La traspuesta de una matriz traspuesta es la matriz original misma. Por tanto, $(\mathbf{A}')' = \mathbf{A}$.
2. Si $\mathbf{A}$ y $\mathbf{B}$ son conformables para la adición, entonces $\mathbf{C} = \mathbf{A} + \mathbf{B}$ y $\mathbf{C}' = (\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$. Es decir, la traspuesta de la suma de dos matrices es la suma de sus traspuestas.
3. Si $\mathbf{AB}$ está definido, entonces $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. Es decir, la traspuesta del producto de dos matrices es el producto de sus traspuestas en orden contrario. Esto puede generalizarse: $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$.
4. La traspuesta de una matriz identidad $\mathbf{I}$ es la matriz identidad misma; es decir $\mathbf{I}' = \mathbf{I}$.
5. La traspuesta de un escalar es el escalar mismo. Por tanto, si $\lambda$ es un escalar, $\lambda' = \lambda$.
6. La traspuesta de $(\lambda\mathbf{A})'$ es $\lambda\mathbf{A}'$, donde $\lambda$ es un escalar. [*Nota:* $(\lambda\mathbf{A})' = \mathbf{A}'\lambda' = \mathbf{A}'\lambda = \lambda\mathbf{A}'$.]
7. Si $\mathbf{A}$ es una matriz cuadrada tal que $\mathbf{A} = \mathbf{A}'$, entonces $\mathbf{A}$ es una matriz simétrica. (Véase la definición de matriz simétrica de la sección B.2.)

### Inversión de matrices

La inversa de una matriz cuadrada $\mathbf{A}$, denotada por $\mathbf{A}^{-1}$ (se lee $\mathbf{A}$ inversa), si existe, es una matriz cuadrada única tal que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

donde $\mathbf{I}$ es una matriz identidad cuyo orden es el mismo que el de $\mathbf{A}$. Por ejemplo,

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} -1 & \frac{1}{2} \\ \frac{6}{8} & -\frac{1}{4} \end{bmatrix} \qquad \mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

Veremos cómo calcular $\mathbf{A}^{-1}$ después de estudiar el tema de determinantes. Mientras tanto observe estas propiedades de la inversa.

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$; es decir, la inversa del producto de dos matrices es el producto de sus inversas en orden opuesto.
2. $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$; es decir, la traspuesta de $\mathbf{A}$ inversa es la inversa de $\mathbf{A}$ traspuesta.

## B.4 Determinantes

Por cada matriz cuadrada $\mathbf{A}$ existe un número (escalar) conocido como el determinante de la matriz, que se denota por det $\mathbf{A}$ o por el símbolo $|\mathbf{A}|$, donde $|\ |$ significa "el determinante de". Observe que una matriz por sí misma no tiene valor numérico, pero el determinante de una matriz es un número.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -7 \\ 2 & 5 & 0 \\ 3 & 8 & 6 \end{bmatrix} \qquad |\mathbf{A}| = \begin{vmatrix} 1 & 3 & -7 \\ 2 & 5 & 0 \\ 3 & 8 & 6 \end{vmatrix}$$

El $|\mathbf{A}|$ en este ejemplo se denomina determinante de orden 3 porque está asociado con una matriz de orden $3 \times 3$.

## Cálculo de un determinante

El proceso de encontrar el valor de un determinante se conoce como *evaluación*, *expansión* o *reducción* del determinante. Esto se logra al manipular los elementos de la matriz en una forma bien definida.

*Evaluación de un determinante de 2 × 2*

Si

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

su determinante se evalúa de la siguiente manera:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

que se obtiene al multiplicar en cruz los elementos de la diagonal principal y restar de ellos la multiplicación en cruz de los elementos de la otra diagonal de la matriz **A**, como indican las flechas.

*Evaluación de un determinante de 3 × 3*

Si

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

entonces

$$|\mathbf{A}| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

Un examen cuidadoso de la evaluación de un determinante de $3 \times 3$ muestra que:

1. Cada término en la expansión del determinante contiene uno y sólo un elemento de cada renglón y de cada columna.
2. El número de elementos en cada término es el mismo que el número de renglones (o de columnas) en la matriz. Por tanto, un determinante de $2 \times 2$ tiene dos elementos en cada término de su expansión, un determinante de $3 \times 3$ tiene tres elementos en cada término de su expansión, y así sucesivamente.
3. Los términos en la expansión alternan su signo de + a −.
4. Un determinante de $2 \times 2$ tiene dos términos en su expansión y un determinante de $3 \times 3$ tiene seis términos en su expansión. La regla general es: El determinante de orden $N \times N$ tiene $N! = N(N-1)(N-2)\cdots 3 \cdot 2 \cdot 1$ términos en su expansión, donde $N!$ se lee "$N$ factorial". Según esta regla, un determinante de orden $5 \times 5$ tendrá $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ términos en su expansión.[1]

## Propiedades de los determinantes

1. Una matriz cuyo determinante tiene un valor de cero se denomina **matriz singular**, mientras que aquélla con un determinante diferente de cero se denomina **matriz no singular**. No hay inversa de una matriz como la recién definida para una matriz singular.

[1] Para evaluar el determinante de una matriz **A** de $N \times N$, consulte las referencias.

2. Si todos los elementos de cualquier renglón de **A** son cero, su determinante es cero. Por tanto,

$$|\mathbf{A}| = \begin{vmatrix} 0 & 0 & 0 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{vmatrix} = 0$$

3. $|\mathbf{A}'| = |\mathbf{A}|$; es decir, los determinantes de **A** y de **A** traspuesta son los mismos.
4. El intercambio de dos renglones cualesquiera o de dos columnas cualesquiera de una matriz **A** cambian el signo de $|\mathbf{A}|$.

**EJEMPLO 1**

Si

$$\mathbf{A} = \begin{bmatrix} 6 & 9 \\ -1 & 4 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} -1 & 4 \\ 6 & 9 \end{bmatrix}$$

donde **B** se obtiene al intercambiar los renglones de **A**, entonces

$$|\mathbf{A}| = 24 - (-9) \quad \text{y} \quad |\mathbf{B}| = -9 - (24)$$
$$= 33 \qquad\qquad\qquad = -33$$

5. Si cada elemento de un renglón o de una columna de **A** se multiplica por un escalar $\lambda$, entonces $|\mathbf{A}|$ se multiplica por $\lambda$.

**EJEMPLO 2**

Si

$$\lambda = 5 \quad \text{y} \quad \mathbf{A} = \begin{bmatrix} 5 & -8 \\ 2 & 4 \end{bmatrix}$$

y multiplicamos el primer renglón de **A** por 5 para obtener

$$\mathbf{B} = \begin{bmatrix} 25 & -40 \\ 2 & 4 \end{bmatrix}$$

se ve que $|\mathbf{A}| = 36$ y $|\mathbf{B}| = 180$, que es $5\,|\mathbf{A}|$.

6. Si dos renglones o columnas de una matriz son idénticas, su determinante es cero.
7. Si un renglón o una columna de una matriz es un múltiplo de otro renglón o columna de esa matriz, su determinante es cero. Por tanto, si

$$\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 2 & 4 \end{bmatrix}$$

donde el primer renglón de **A** es el doble de su segundo renglón, $|\mathbf{A}| = 0$. De forma más general, si cualquier renglón (columna) de una matriz es una combinación lineal de otros renglones (columnas), su determinante es cero.

8. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$; es decir, el determinante del producto de dos matrices es el producto de sus determinantes (individuales).

## Rango de una matriz

El rango de una matriz es el orden de la submatriz cuadrada más grande cuyo determinante no sea cero.

**EJEMPLO 3**

$$\mathbf{A} = \begin{bmatrix} 3 & 6 & 6 \\ 0 & 4 & 5 \\ 3 & 2 & 1 \end{bmatrix}$$

Puede verse que $|\mathbf{A}| = 0$. En otras palabras, **A** es una matriz singular. Por tanto, aunque su orden es $3 \times 3$, su rango es menor que 3. En realidad, es 2, pues se puede encontrar una submatriz $2 \times 2$ cuyo determinante no es cero. Por ejemplo, si borramos el primer renglón y la primera columna de **A**, obtenemos

$$\mathbf{B} = \begin{bmatrix} 4 & 5 \\ 2 & 1 \end{bmatrix}$$

cuyo determinante es $-6$, que es diferente de cero. Así, el rango de **A** es 2. Como ya mencionamos, la inversa de una matriz singular no existe. Por consiguiente, para una matriz **A** de $N \times N$, su rango debe ser $N$ para que su inversa exista; si es menor que $N$, **A** es singular.

## Menor

Si se borra el renglón $i$-ésimo y la columna $j$-ésima de una matriz **A** de $N \times N$, el determinante de la submatriz resultante se denomina el **menor** del elemento $a_{ij}$ (el elemento en el intercepto del renglón $i$-ésimo y de la columna $j$-ésima) y se denota por $|\mathbf{M}_{ij}|$.

**EJEMPLO 4**

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

El menor de $a_{11}$ es

$$|\mathbf{M}_{11}| = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32}$$

En forma similar, el menor de $a_{21}$ es

$$|\mathbf{M}_{21}| = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = a_{12}a_{33} - a_{13}a_{32}$$

De la misma manera, pueden encontrarse menores de otros elementos de **A**.

## Cofactor

El cofactor del elemento $a_{ij}$ de una matriz **A** de $N \times N$, denotado por $c_{ij}$, se define como

$$c_{ij} = (-1)^{i+j} |\mathbf{M_{ij}}|$$

En otras palabras, un cofactor es un menor *con un signo asociado*, con signo positivo si $i + j$ es par y negativo si $i + j$ es impar. Por tanto, el cofactor del elemento $a_{11}$ de la matriz **A** de $3 \times 3$ dado antes es $a_{22}a_{33} - a_{23}a_{32}$, mientras que el cofactor del elemento $a_{21}$ es $-(a_{12}a_{33} - a_{13}a_{32})$ porque la suma de los subíndices 2 y 1 es 3, un número impar.

### *Matriz de cofactores*

Al remplazar los elementos $a_{ij}$ de una matriz **A** por sus cofactores obtenemos una matriz conocida como **matriz de cofactores** de **A**, denotada por (cof **A**).

### *Matriz adjunta*

La matriz adjunta, escrita como (adj **A**), es la traspuesta de la matriz de cofactores; es decir, (adj **A**) = (cof **A**)′.

## B.5 Forma de encontrar la inversa de una matriz cuadrada

Si **A** es cuadrada y no singular (es decir, $|\mathbf{A}| \neq 0$), su inversa $\mathbf{A}^{-1}$ se encuentra de la siguiente manera:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}(\text{adj } \mathbf{A})$$

Los pasos comprendidos en el cálculo son los siguientes:

1. Encontrar el determinante de **A**. Si es diferente de cero, proceda al paso 2.
2. Remplazar cada elemento $a_{ij}$ de **A** por su cofactor para obtener la matriz de cofactores.
3. Trasponer la matriz de cofactores para obtener la matriz adjunta.
4. Dividir cada elemento de la matriz adjunta por $|\mathbf{A}|$.

**EJEMPLO 5**

Encuentre la inversa de la matriz

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 7 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

**Paso 1.** Primero se encuentra el determinante de la matriz. Al aplicar las reglas de expansión del determinante de $3 \times 3$ dado antes, obtenemos $|\mathbf{A}| = -24$.

**Paso 2.** Obtenemos ahora la matriz de cofactores, por ejemplo, **C**:

$$\mathbf{C} = \begin{bmatrix} \begin{vmatrix} 7 & 4 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 5 & 4 \\ 2 & 3 \end{vmatrix} & \begin{vmatrix} 5 & 7 \\ 2 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 7 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 5 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 5 & 7 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 17 & -7 & -9 \\ -3 & -3 & 3 \\ -13 & 11 & -3 \end{bmatrix}$$

**Paso 3.** Al trasponer la matriz de cofactores anterior, obtenemos la siguiente matriz adjunta:

$$(\text{adj } \mathbf{A}) = \begin{bmatrix} 17 & -3 & -13 \\ -7 & -3 & 11 \\ -9 & 3 & -3 \end{bmatrix}$$

**Paso 4.** Ahora dividimos los elementos de (adj **A**) entre el valor del determinante −24 para obtener

$$\mathbf{A}^{-1} = -\frac{1}{24}\begin{bmatrix} 17 & -3 & -13 \\ -7 & -3 & 11 \\ -9 & 3 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{17}{24} & \frac{3}{24} & \frac{13}{24} \\ \frac{7}{24} & \frac{3}{24} & -\frac{11}{24} \\ \frac{9}{24} & -\frac{3}{24} & \frac{3}{24} \end{bmatrix}$$

Se verifica fácilmente que

$$\mathbf{AA}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

que es una matriz identidad. El lector debe corroborar que para el ejemplo ilustrativo del apéndice C, la inversa de la matriz $\mathbf{X}'\mathbf{X}$ es la que aparece en la ecuación (C.10.5).

## B.6 Diferenciación matricial

Para seguir el material del apéndice CA, sección CA.2, necesitamos algunas reglas respecto de la diferenciación de matrices.

**REGLA 1**

Si $\mathbf{a}' = [a_1 \quad a_2 \ldots a_n]$ es un vector renglón de números y

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

es un vector columna de las variables $x_1, x_2, \ldots, x_n$, entonces

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial\mathbf{x}} = \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

**REGLA 2**

Considere la matriz $\mathbf{x}'\mathbf{A}\mathbf{x}$ tal que

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Entonces,

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

que es un vector columna de $n$ elementos, o

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = 2\mathbf{x}'\mathbf{A}$$

que es un vector renglón de $n$ elementos.

## Referencias


Chiang, Alpha C., *Fundamental Methods of Mathematical Economics*, 3a. ed., McGraw-Hill, Nueva York, 1984, capítulos 4 y 5. Allí se presenta un análisis elemental.

Hadley, G., *Linear Algebra*, Addison-Wesley, Reading, Massachusetts, 1961. Es un análisis avanzado.

# Apéndice C

# Método matricial para el modelo de regresión lineal

En este apéndice presentamos el modelo clásico de regresión lineal de $k$ variables ($Y$ y $X_2, X_3, \ldots, X_k$) en notación de álgebra matricial. En teoría, el modelo de $k$ variables es una extensión lógica de los modelos de dos y tres variables considerados hasta ahora en este texto. Por consiguiente, en este capítulo veremos muy pocos conceptos nuevos, a excepción de la notación matricial.[1]

Una gran ventaja del álgebra matricial sobre la escalar (álgebra elemental que trata con escalares o números reales) es que proporciona un método compacto de manejo de modelos de regresión que implican cualquier número de variables; una vez formulado y resuelto en notación matricial el modelo de $k$ variables, la solución es aplicable a una, dos, tres o cualquier número de variables.

## C.1 Modelo de regresión lineal con $k$ variables

Si generalizamos los modelos de regresión lineal de dos y tres variables, el modelo de regresión poblacional de $k$ variables (FRP) con la variable dependiente $Y$ y $k-1$ variables explicativas $X_2, X_3, \ldots, X_k$ puede escribirse así:

$$\text{FRP:}\quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \qquad i = 1, 2, 3, \ldots, n \tag{C.1.1}$$

donde $\beta_1$ = el intercepto, $\beta_2$ a $\beta_k$ = coeficientes parciales de pendientes, $u$ = término de perturbación estocástica e $i$ = $i$-ésima observación, con $n$ como tamaño de la población. La FRP (C.1.1) se interpreta en la forma usual: la media o el valor esperado de $Y$ condicionado a los valores fijos (en muestreo repetido) de $X_2, X_3, \ldots, X_k$, es decir, $E(Y \mid X_{2i}, X_{3i}, \ldots, X_{ki})$.

[1] Los lectores que no estén familiarizados con el álgebra matricial deben revisar el **apéndice B** antes de continuar. El **apéndice B** proporciona las bases del álgebra matricial requeridas para comprender este capítulo.

La ecuación (C.1.1) es una expresión abreviada para el siguiente conjunto de $n$ ecuaciones simultáneas:

$$
\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + u_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + u_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_k X_{kn} + u_n
\end{aligned}
\tag{C.1.2}
$$

El sistema de ecuaciones (C.1.2) se escribe en una forma alterna aunque más ilustrativa:[2]

$$
\underset{n \times 1}{\underset{\mathbf{y}}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}} = \underset{n \times k}{\underset{\mathbf{X}}{\begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}}} \underset{k \times 1}{\underset{\boldsymbol{\beta}}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}} + \underset{n \times 1}{\underset{\mathbf{u}}{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}}}
\tag{C.1.3}
$$

donde $\mathbf{y}$ = vector columna $n \times 1$ de observaciones sobre la variable dependiente $Y$
$\mathbf{X}$ = matriz $n \times k$, con $n$ observaciones sobre las $k - 1$ variables $X_2$ a $X_k$, y la primera columna de números 1 representa el término del intercepto. (Esta matriz se conoce también como **matriz de datos**.)
$\boldsymbol{\beta}$ = vector columna $k \times 1$ de los parámetros desconocidos $\beta_1, \beta_2, \ldots, \beta_k$
$\mathbf{u}$ = vector columna $n \times 1$ de $n$ perturbaciones $u_i$

Con las reglas de multiplicación y adición de matrices, el lector debe verificar que los sistemas (C.1.2) y (C.1.3) sean equivalentes.

El sistema (C.1.3) se conoce como *representación matricial del modelo de regresión lineal general (de k variables)*. Se escribe en forma más compacta como

$$
\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \; \underset{k \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\mathbf{u}}
\tag{C.1.4}
$$

Donde no haya confusión sobre las dimensiones u órdenes de la matriz **X** y de los vectores **y**, **β** y **u**, la ecuación (C.1.4) se escribe tan sólo como:

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}
\tag{C.1.5}
$$

Como ilustración de la representación matricial, considere el modelo de dos variables consumo-ingreso del capítulo 3, a saber: $Y_i = \beta_1 + \beta_2 X_i + u_i$, donde $Y$ es el gasto de consumo y $X$ es el ingreso.

[2] Según la notación del **apéndice B**, los vectores se representan con letras minúsculas en negritas y las matrices con letras mayúsculas en negritas.

Con la información de la tabla 3.2 podemos escribir la formulación matricial así:

$$\begin{bmatrix} 70 \\ 65 \\ 90 \\ 95 \\ 110 \\ 115 \\ 120 \\ 140 \\ 155 \\ 150 \end{bmatrix} = \begin{bmatrix} 1 & 80 \\ 1 & 100 \\ 1 & 120 \\ 1 & 140 \\ 1 & 160 \\ 1 & 180 \\ 1 & 200 \\ 1 & 220 \\ 1 & 240 \\ 1 & 260 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{bmatrix} \tag{C.1.6}$$

$$\underset{10 \times 1}{\mathbf{y}} = \underset{10 \times 2}{\mathbf{X}} \ \underset{2 \times 1}{\boldsymbol{\beta}} + \underset{10 \times 1}{\mathbf{u}}$$

Como en los casos de dos y tres variables, el objetivo es estimar los parámetros de la regresión múltiple (C.1.1) y efectuar inferencias sobre ellos a partir de la información disponible. En la notación matricial esto equivale a estimar $\boldsymbol{\beta}$ y a inferir sobre él. Para fines de estimación, podemos utilizar el método de mínimos cuadrados ordinarios (MCO) o el método de máxima verosimilitud (MV). Pero, como ya mencionamos, estos dos métodos producen valores estimados idénticos de los coeficientes de regresión.[3] Por consiguiente, limitaremos nuestra atención al método de MCO.

## C.2 Supuestos del modelo clásico de regresión lineal en notación matricial

Los supuestos en los cuales se basa el modelo clásico de regresión lineal están en la tabla C.1; se presentan en notación escalar y en notación matricial. El supuesto 1 de (C.2.1) significa que el valor esperado del vector de perturbaciones **u**, es decir, de cada uno de sus elementos, es cero. Más explícitamente, $E(\mathbf{u}) = \mathbf{0}$ significa

$$E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{C.2.1}$$

El supuesto 2 [ecuación (C.2.2)] es una forma compacta de expresar los dos supuestos de (3.2.5) y (3.2.2) con notación escalar. Para ver esto, escribimos

$$E(\mathbf{u}\mathbf{u}') = E \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} [u_1 \quad u_2 \quad \cdots \quad u_n]$$

[3] La prueba de que esto es así en el caso de *k* variables se encuentra en las notas del capítulo 4.

**TABLA C.1**
**Supuestos del modelo clásico de regresión lineal**

| Notación escalar | | | Notación matricial |
|---|---|---|---|
| 1. $E(u_i) = 0$, | para cada $i$ | (3.2.1) | 1. $E(\mathbf{u}) = \mathbf{0}$ donde **u** y **0** son vectores columna $n \times 1$, con **0** como vector nulo |
| 2. $E(u_i, u_j) = 0$<br>$= \sigma^2$ | $i \neq j$<br>$i \neq j$ | (3.2.5)<br>(3.2.2) | 2. $E(\mathbf{uu}') = \sigma^2\mathbf{I}$ donde **I** es una matriz de identidad $n \times n$ |
| 3. $X_2, X_3, \ldots, X_k$ son fijas o no estocásticas | | | 3. La matriz **X**, $n \times k$ es no estocástica; es decir, consiste en un conjunto de números fijos |
| 4. No hay relación lineal exacta entre las variables $X$; es decir, no hay multicolinealidad | | (7.1.9) | 4. El rango de **X** es $p(\mathbf{X}) = k$, donde $k$ es el número de columnas en **X** y $k$ es menor que el número de observaciones, $n$ |
| 5. Para las pruebas de hipótesis $u_i \sim N(0, \sigma^2)$ | | (4.2.4) | 5. El vector **u** tiene una distribución normal multivariada, es decir, $\mathbf{u} \sim N(0, \sigma^2\mathbf{I})$ |

donde $\mathbf{u}'$ es la traspuesta del vector columna **u**, o vector renglón. Al efectuar la multiplicación, obtenemos

$$E(\mathbf{uu}') = E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \cdots & \cdots & \cdots & \cdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

Al aplicar el operador de valor esperado $E$ a cada elemento de la matriz anterior, obtenemos

$$E(\mathbf{uu}') = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} \tag{C.2.2}$$

Debido a los supuestos de homoscedasticidad y de no correlación serial, la matriz (C.2.2) se reduce a

$$E(\mathbf{uu}') = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{C.2.3}$$

$$= \sigma^2\mathbf{I}$$

donde **I** es una matriz identidad $n \times n$.

La matriz (C.2.2) [y su representación en (C.2.3)] se denomina **matriz de varianza-covarianza** de las perturbaciones $u_i$; los elementos sobre la diagonal principal (que van de la esquina superior izquierda a la esquina inferior derecha) de esta matriz dan las varianzas, y los elementos

por fuera de la diagonal principal dan las covarianzas.[4] Observe que la matriz de varianza-covarianza es **simétrica**: los elementos por encima y por debajo de la diagonal principal son reflejos unos de los otros.

El *supuesto 3* de la tabla C.1 establece que la matriz **X** de $n \times k$ es no estocástica; es decir, consta de números fijos. Como ya dijimos, el análisis de regresión es de regresión condicional, es decir, condicional a los valores fijos de las variables $X$.

El *supuesto 4* establece que la matriz **X** tiene rango columna completo igual a $k$, el número de columnas en la matriz. Esto significa que las columnas de la matriz $X$ son linealmente independientes; es decir, no hay **relación lineal exacta** entre las variables $X$. En otras palabras, no hay multicolinealidad. En notación escalar esto equivale a decir que no existe un conjunto de números $\lambda_1, \lambda_2, \ldots, \lambda_k = 0$ no todos iguales a cero tales que [cf. (7.1.8)]

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0 \qquad \text{(C.2.4)}$$

donde $X_{1i} = 1$ para todo $i$ (para dar cabida a la columna de números 1 en la matriz **X**). En notación matricial, (C.2.4) se representa como

$$\boldsymbol{\lambda}'\mathbf{x} = 0 \qquad \text{(C.2.5)}$$

donde $\lambda'$ es un vector renglón de $1 \times k$ y **x** es un vector columna de $k \times 1$.

Si existe una relación lineal exacta como (C.2.4), se dice que las variables son colineales. Si, por otra parte, (C.2.4) se cumple sólo si $\lambda_1 = \lambda_2 = \lambda_3 = \cdots = 0$, entonces se dice que las variables $X$ son linealmente independientes. Dimos una razón intuitiva para el supuesto de *no multicolinealidad* en el capítulo 7, y analizamos más a fondo este supuesto en el capítulo 10.

## C.3 Estimación por MCO

Para obtener la estimación por MCO de $\boldsymbol{\beta}$, primero escribimos la regresión muestral de $k$-variables (FRM):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \qquad \text{(C.3.1)}$$

la cual se escribe en forma más compacta en notación matricial como:

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} \qquad \text{(C.3.2)}$$

y en forma matricial como

$$\underset{n \times 1}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n \times k}{\begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}} \underset{k \times 1}{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}} + \underset{n \times 1}{\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}} \qquad \text{(C.3.3)}$$

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \quad \underset{k \times 1}{\hat{\boldsymbol{\beta}}} + \underset{n \times 1}{\hat{\mathbf{u}}}$$

donde $\hat{\boldsymbol{\beta}}$ es un vector columna de $k$ elementos compuesto por los estimadores de MCO de los coeficientes de regresión, y donde $\hat{\mathbf{u}}$ es un vector columna de $n \times 1$ con $n$ residuos.

[4] Por definición, la varianza de $u_i = E[u_i - E(u_i)]^2$ y la covarianza entre $u_i$ y $u_j = E[u_i - E(u_i)][u_j - E(u_j)]$. Pero, debido al supuesto $E(u_i) = 0$ para cada $i$, tenemos la matriz de varianza-covarianza (C.2.3).

Como en los modelos de dos y tres variables, en el caso de $k$ variables los estimadores de MCO se obtienen al reducir

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2 \tag{C.3.4}$$

donde $\sum \hat{u}_i^2$ es la suma de cuadrados residual (SCR). En notación matricial, esto equivale a reducir $\hat{\mathbf{u}}'\hat{\mathbf{u}}$, pues

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = [\hat{u}_1 \quad \hat{u}_2 \quad \cdots \quad \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2 = \sum \hat{u}_i^2 \tag{C.3.5}$$

Ahora, de (C.3.2) obtenemos

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \tag{C.3.6}$$

Por consiguiente,

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{C.3.7}$$

donde aprovechamos las propiedades de la traspuesta de una matriz, a saber: $(\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'$; y, como $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ es un escalar (un número real), es igual a su traspuesta $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$.

La ecuación (C.3.7) es la representación matricial de (C.3.4). En notación escalar, el método de MCO consiste en estimar $\beta_1, \beta_2, \ldots, \beta_k$ de manera que $\sum \hat{u}_i^2$ sea lo más pequeño posible. Esto se logra al diferenciar parcialmente (C.3.4) respecto de $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ e igualar a cero las expresiones resultantes. Este proceso produce $k$ ecuaciones simultáneas con $k$ incógnitas, que son las ecuaciones normales de la teoría de mínimos cuadrados. Como se ve en el apéndice CA, sección CA.1, estas ecuaciones son las siguientes:

$$\begin{aligned} n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\ \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i}X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i}X_{ki} &= \sum X_{2i}Y_i \\ \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{3i}X_{2i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i}X_{ki} &= \sum X_{3i}Y_i \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\cdots\cdots\cdots \\ \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki}X_{2i} + \hat{\beta}_3 \sum X_{ki}X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki}Y_i \end{aligned}$$

(C.3.8)[5]

En forma matricial, la ecuación (C.3.8) se representa de la siguiente manera:

$$\underset{(\mathbf{X}'\mathbf{X})}{\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \cdots & \sum X_{ki}^2 \end{bmatrix}} \underset{\hat{\boldsymbol{\beta}}}{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}} = \underset{\mathbf{X}'}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{32} & \cdots & X_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix}} \underset{\mathbf{y}}{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}} \tag{C.3.9}$$

[5] Estas ecuaciones pueden recordarse fácilmente. Empiece con la ecuación $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$. Al sumar esta ecuación sobre los $n$ valores se obtiene la primera ecuación de (C.3.8); al multiplicar ésta por $X_2$ en ambos lados y sumar sobre $n$ se obtiene la segunda ecuación; multiplique ésta por $X_3$ en ambos lados y sume sobre $n$ para obtener la tercera ecuación, y así sucesivamente. A propósito, observe que a partir de la primera ecuación en (C.3.8) resulta $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k$ [consulte (7.4.6)].

o, en forma más compacta, como

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{C.3.10}$$

Note las siguientes características de la matriz $(\mathbf{X}'\mathbf{X})$: 1) Proporciona las sumas simples de cuadrados y productos cruzados de las variables $X$, una de las cuales es el término del intercepto que toma el valor 1 para cada observación. Los elementos sobre la diagonal principal dan las sumas simples de cuadrados y los que no están en la diagonal principal dan las sumas simples de productos cruzados (por *simples* nos referimos a que están expresadas en las unidades originales de medición). 2) Es simétrica, pues el producto cruzado entre $X_{2i}$ y $X_{3i}$ es el mismo que entre $X_{3i}$ y $X_{2i}$. 3) Es de orden $(k \times k)$, es decir, tiene $k$ renglones y $k$ columnas.

En (C.3.10), las cantidades conocidas son $(\mathbf{X}'\mathbf{X})$ y $(\mathbf{X}'\mathbf{y})$ (el producto cruzado entre las variables $X$ y $y$) y la incógnita es $\hat{\boldsymbol{\beta}}$. Ahora, mediante álgebra matricial, si existe la inversa de $(\mathbf{X}'\mathbf{X})$, es decir, $(\mathbf{X}'\mathbf{X})^{-1}$, entonces premultiplicamos ambos lados de (C.3.10) por esta inversa para obtener

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Pero, como $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}$ es una matriz identidad de orden $k \times k$, obtenemos

$$\mathbf{I}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

o

$$\underset{k\times 1}{\hat{\boldsymbol{\beta}}} = \underset{k\times k}{(\mathbf{X}'\mathbf{X})^{-1}}\ \underset{(k\times n)}{\mathbf{X}'}\ \underset{(n\times 1)}{\mathbf{y}} \tag{C.3.11}$$

**La ecuación (C.3.11) es un resultado fundamental de la teoría de MCO en notación matricial**. Muestra cómo se estima el vector $\hat{\boldsymbol{\beta}}$ a partir de la información dada. Aunque (C.3.11) se obtuvo de (C.3.9), se puede obtener directamente de (C.3.7) al diferenciar $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ respecto de $\hat{\boldsymbol{\beta}}$. La prueba se da en el apéndice CA, sección CA.2.

## Una ilustración

Como ilustración de los métodos matriciales desarrollados hasta el momento, analicemos de nuevo el ejemplo de consumo-ingreso con los datos de la ecuación (C.1.6). Para el caso de dos variables tenemos

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ & \cdots \\ 1 & X_N \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

y

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Con la información de (C.1.6), obtenemos

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 10 & 1\,700 \\ 1\,700 & 322\,000 \end{bmatrix}$$

y

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1\,110 \\ 205\,500 \end{bmatrix}$$

Recurrimos a las reglas de inversión de matriz del **apéndice B**, **sección B.3**, y vemos que la inversa de la matriz (**X′X**) anterior es

$$\mathbf{X}'\mathbf{X}^{-1} = \begin{bmatrix} 0.97576 & -0.005152 \\ -0.005152 & 0.0000303 \end{bmatrix}$$

Por consiguiente,

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0.97576 & -0.005152 \\ -0.005152 & 0.0000303 \end{bmatrix} \begin{bmatrix} 1\,110 \\ 205\,500 \end{bmatrix}$$

$$= \begin{bmatrix} 24.4545 \\ 0.5079 \end{bmatrix}$$

Con la computadora obtuvimos $\hat{\beta}_1 = 24.4545$ y $\hat{\beta}_2 = 0.5091$. La diferencia entre las dos estimaciones se debe a errores de redondeo. A propósito, observe que al trabajar con calculadora de escritorio, es esencial obtener resultados con un número significativo de dígitos para reducir los errores de redondeo.

## Matriz de varianza-covarianza de $\hat{\boldsymbol{\beta}}$

Los métodos matriciales permiten desarrollar fórmulas no sólo para la varianza de $\hat{\beta}_i$, cualquier elemento dado de $\hat{\boldsymbol{\beta}}$, sino también para la covarianza entre dos elementos de $\hat{\boldsymbol{\beta}}$ cualesquiera, por ejemplo, $\hat{\beta}_i$ y $\hat{\beta}_j$. Se necesitan estas varianzas y covarianzas para fines de inferencia estadística.

Por definición, la matriz de varianza-covarianza de $\hat{\boldsymbol{\beta}}$ es [consulte (C.2.2)]

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\}$$

la cual se escribe explícitamente como

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \text{var}\,(\hat{\beta}_1) & \text{cov}\,(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \text{cov}\,(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}\,(\hat{\beta}_2, \hat{\beta}_1) & \text{var}\,(\hat{\beta}_2) & \cdots & \text{cov}\,(\hat{\beta}_2, \hat{\beta}_k) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}\,(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}\,(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \text{var}\,(\hat{\beta}_k) \end{bmatrix} \tag{C.3.12}$$

Se muestra en el apéndice CA, sección CA.3, que la matriz anterior de varianza-covarianza se obtiene a partir de la siguiente fórmula:

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{C.3.13}$$

donde $\sigma^2$ es la varianza homoscedástica de $u_i$ y $(\mathbf{X}'\mathbf{X})^{-1}$ es la matriz inversa que aparece en la ecuación (C.3.11), la cual da el estimador de MCO, $\hat{\boldsymbol{\beta}}$.

En los modelos de regresión lineal con dos y tres variables, un estimador insesgado de $\sigma^2$ estaba dado por $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$ y $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$, respectivamente. En el caso de $k$ variables, la fórmula correspondiente es

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum \hat{u}_i^2}{n-k}\\ &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\end{aligned} \tag{C.3.14}$$

donde ahora hay $n-k$ gl. (¿Por qué?)

Aunque, en principio, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ se calcula a partir de los residuos estimados, en la práctica se obtiene directamente de la siguiente manera. Recuerde que $\sum \hat{u}_i^2$ (= SCR) = SCT − SCE, y en el caso de dos variables podemos escribir

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

y en el caso de tres variables

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

Al extender este principio, se ve que para el modelo de $k$ variables

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \cdots - \hat{\beta}_k \sum y_i x_{ki} \tag{C.3.15}$$

En notación matricial,

$$\text{SCT: } \sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.16}$$

$$\text{SCE: } \hat{\beta}_2 \sum y_i x_{2i} + \cdots + \hat{\beta}_k \sum y_i x_{ki} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.17}$$

donde el término $\bar{Y}^2$ se conoce como corrección para la media.[6] Por consiguiente,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{C.3.18}$$

Una vez obtenida $\hat{\mathbf{u}}'\hat{\mathbf{u}}$, $\hat{\sigma}^2$ se calcula fácilmente de (C.3.14), lo cual, a su vez, permite estimar la matriz de varianza-covarianza (C.3.13).

Para nuestro ejemplo ilustrativo,

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= 132\,100 - [24.4545 \quad 0.5091]\begin{bmatrix} 1\,110 \\ 205\,500 \end{bmatrix}\\ &= 337.373\end{aligned}$$

Por tanto, $\hat{\sigma}^2 = (337.273/8) = 42.1591$, que es aproximadamente el valor obtenido ya en el capítulo 3.

[6] *Nota:* $\sum y_i^2 = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2$. Por consiguiente, sin el término de corrección, **y′y** dará tan sólo la suma sencilla de cuadrados, no la suma de las desviaciones al cuadrado.

### Propiedades del vector de MCO $\hat{\boldsymbol{\beta}}$

En los casos de dos y tres variables, sabemos que los estimadores de MCO son lineales e insesgados, y en la clase de todos los estimadores lineales e insesgados, éstos tienen varianza mínima (propiedad de Gauss-Markov). En resumen, los estimadores de MCO son los mejores estimadores lineales insesgados (MELI). Esta propiedad se extiende a todo el vector $\hat{\boldsymbol{\beta}}$; es decir, $\hat{\boldsymbol{\beta}}$ es lineal (cada uno de sus elementos es una función lineal de $Y$, la variable dependiente). $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, es decir, el valor esperado de cada elemento de $\hat{\boldsymbol{\beta}}$ es igual al elemento correspondiente de la verdadera $\boldsymbol{\beta}$, y en la clase de todos los estimadores lineales e insesgados de $\boldsymbol{\beta}$, el estimador de MCO, $\hat{\boldsymbol{\beta}}$ tiene varianza mínima.

La prueba se da en el apéndice CA, sección CA.4. Como establecimos en la introducción, el caso de $k$ variables es, en su mayoría, una extensión directa de los casos de dos y tres variables.

## C.4 Coeficiente de determinación $R^2$ en notación matricial

El coeficiente de determinación $R^2$ se ha definido como

$$R^2 = \frac{\text{SCE}}{\text{SCT}}$$

En el caso de dos variables,

$$R^2 = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} \tag{3.5.6}$$

y en el caso de tres variables

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \tag{7.5.5}$$

Si generalizamos, para el caso de $k$ variables obtenemos

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2} \tag{C.4.1}$$

Con (C.3.16) y (C.3.17), la ecuación (C.4.1) se escribe como

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \tag{C.4.2}$$

lo cual da la representación matricial de $R^2$.

Para nuestro ejemplo ilustrativo,

$$\begin{aligned}\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} &= [24.3571 \quad 0.5079]\begin{bmatrix} 1\ 110 \\ 205\ 500 \end{bmatrix} \\ &= 131\ 409.831 \\ \mathbf{y}'\mathbf{y} &= 132\ 100\end{aligned}$$

y

$$n\bar{Y}^2 = 123\ 210$$

Al reemplazar estos valores en (C.4.2) vemos que $R^2 = 0.9224$, que se acerca al valor obtenido antes, salvo por errores de redondeo.

## C.5 Matriz de correlación

En los capítulos anteriores encontramos los coeficientes de correlación de orden cero o simple, $r_{12}$, $r_{13}$, $r_{23}$, y las correlaciones parciales o de primer orden, $r_{12.3}$, $r_{13.2}$, $r_{23.1}$, y sus interrelaciones. En el caso de $k$ variables tendremos en total $k(k-1)/2$ coeficientes de correlación de orden cero. (¿Por qué?) Estas $k(k-1)/2$ correlaciones se sitúan en una matriz, denominada **matriz de correlación R**, de la siguiente manera:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & r_{kk} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix} \tag{C.5.1}$$

donde el subíndice 1, como antes, denota la variable dependiente $Y$ ($r_{12}$ significa el coeficiente de correlación entre $Y$ y $X_2$, y así sucesivamente), y donde se aplica el hecho de que el coeficiente de correlación de una variable respecto de ella misma es siempre 1 ($r_{11} = r_{22} = \cdots = r_{kk} = 1$).

A partir de la matriz de correlación $\boldsymbol{R}$ se obtienen los coeficientes de correlación de primer orden (véase el capítulo 7) y de órdenes superiores como $r_{12.34\ldots k}$. (ejercicio C.4.) Muchos programas de computadora calculan mediante rutinas la matriz $\boldsymbol{R}$. Utilizamos la matriz de correlación en el capítulo 10.

## C.6 Pruebas de hipótesis sobre coeficientes de regresión individuales en notación matricial

Por las razones expresadas en los capítulos anteriores, si el objetivo es tanto la inferencia como la estimación, tenemos que suponer que las perturbaciones $u_i$ siguen alguna distribución de probabilidad. Además, por las razones ya citadas, en el análisis de regresión suele suponerse que cada $u_i$ sigue una distribución normal con media cero y varianza constante $\sigma^2$. En notación matricial, tenemos

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{C.6.1}$$

donde **u** y **0** son vectores columna $n \times 1$ e **I** es una matriz identidad ($n \times n$), con **0** como **vector nulo**.

Con el supuesto de normalidad, sabemos que en los modelos de regresión lineal de dos y tres variables (1) el estimador de MCO $\hat{\beta}_i$, y el estimador de MV $\tilde{\beta}_i$ son idénticos, pero el estimador de MV $\sim \sigma^2$ es sesgado, aunque este sesgo se elimina mediante el estimador de MCO insesgado $\hat{\sigma}^2$; y (2) los estimadores de MCO $\hat{\beta}_i$, también están normalmente distribuidos. Para generalizar, en el caso de $k$ variables es posible demostrar que

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}] \tag{C.6.2}$$

es decir, cada elemento de $\hat{\boldsymbol{\beta}}$ está normalmente distribuido con media igual al elemento correspondiente del verdadero $\boldsymbol{\beta}$ y la varianza está dada por $\sigma^2$ veces el elemento correspondiente de la diagonal de la matriz inversa $(\mathbf{X}'\mathbf{X})^{-1}$.

Como en la práctica se desconoce $\sigma^2$, se estima mediante $\hat{\sigma}^2$. Entonces, por el cambio usual a la distribución *t*, se cumple que cada elemento de $\hat{\boldsymbol{\beta}}$ sigue la distribución *t* con $n - k$ gl. Simbólicamente,

$$t = \frac{\hat{\beta}_i - \beta_i}{\text{ee}\,(\hat{\beta}_i)} \tag{C.6.3}$$

con $n - k$ gl, donde $\hat{\beta}_i$ es cualquier elemento de $\hat{\boldsymbol{\beta}}$.

Por consiguiente, la distribución *t* sirve para probar hipótesis sobre el verdadero $\beta_i$ y establecer intervalos de confianza sobre él. Ya ilustramos el verdadero mecanismo en los capítulos 5 y 8. Para ver un ejemplo completo, consulte la sección C.10.

## C.7 Prueba de significancia global de la regresión: análisis de varianza en notación matricial

En el capítulo 8 desarrollamos la técnica ANOVA 1) para probar la significancia global de la regresión estimada, es decir, para probar la hipótesis nula de que los verdaderos coeficientes de pendiente (parciales) son simultáneamente iguales a cero, y 2) para evaluar la contribución incremental de una variable explicativa. La técnica ANOVA se amplía fácilmente al caso de *k* variables. Recuerde que la técnica ANOVA consiste en descomponer la SCT en dos componentes: la SCE y la SCR. Ya dimos las expresiones matriciales para estas tres sumas de cuadrados en (C.3.16), (C.3.17) y (C.3.18), respectivamente. Los grados de libertad asociados con estas sumas de cuadrados son $n - 1$, $k - 1$ y $n - k$, respectivamente. (¿Por qué?) Entonces, según el capítulo 8, tabla 8.1, podemos elaborar la tabla C.2.

Si suponemos que las perturbaciones $u_i$ están normalmente distribuidas y la hipótesis nula es $\beta_2 = \beta_3 = \cdots = \beta_k = 0$ y, de acuerdo con el capítulo 8, se demuestra que

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2)/(k-1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k)} \tag{C.7.1}$$

sigue la distribución *F* con $k - 1$ y $n - k$ gl.

En el capítulo 8 vimos que, según los supuestos postulados antes, existe una estrecha relación entre *F* y $R^2$, a saber:

$$F = \frac{R^2)(k-1)}{(1-R^2)/(n-k)} \tag{8.4.11}$$

Por consiguiente, la tabla C.2 ANOVA se expresa como tabla C.3. Una ventaja de la tabla C.3 sobre la tabla C.2 es que la totalidad del análisis se realiza en términos de $R^2$; no se requiere considerar el término $(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$, pues éste se cancela en la razón *F*.

**TABLA C.2**
**Formulación matricial de la tabla ANOVA para el modelo de regresión lineal con *k* variables**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| Debido a la regresión (es decir, debido a $X_2, X_3, \ldots, X_k$) | $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$ | $k - 1$ | $\dfrac{\hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{k-1}$ |
| Debido a los residuos | $\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ | $n - k$ | $\dfrac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{n-k}$ |
| Total | $\mathbf{y}'\mathbf{y} - n\bar{Y}^2$ | $n - 1$ | |

**TABLA C.3**
**Tabla ANOVA de $k$ variables en forma matricial en términos de $R^2$**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| Debido a la regresión (es decir, debido a $X_2, X_3, \ldots, X_k$) | $R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$ | $k-1$ | $\dfrac{R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{k-1}$ |
| Debido a los residuos | $(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$ | $n-k$ | $\dfrac{(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{n-k}$ |
| Total | $\mathbf{y}'\mathbf{y} - n\bar{Y}^2$ | $n-1$ | |

## C.8 Pruebas de restricciones lineales: prueba *F* general con notación matricial

En la sección 8.6 presentamos la prueba general $F$ para verificar la validez de las restricciones lineales impuestas sobre uno o más parámetros del modelo de regresión lineal de $k$ variables. Proporcionamos la prueba correspondiente en (8.6.9) [o su equivalente, ecuación (8.6.10)]. La matriz correspondiente a (8.6.9) se deriva fácilmente.

Sea

$$\hat{\mathbf{u}}_R = \text{vector residuo de la regresión de mínimos cuadrados restringidos}$$
$$\hat{\mathbf{u}}_{NR} = \text{vector residuo de la regresión de mínimos cuadrados no restringidos}$$

Entonces

$$\hat{\mathbf{u}}_R'\hat{\mathbf{u}}_R = \sum \hat{u}_R^2 = \text{SCR de la región restringida}$$
$$\hat{\mathbf{u}}_{NR}'\hat{\mathbf{u}}_{NR} = \sum \hat{u}_{NR}^2 = \text{SCR de la región no restringida}$$
$$m = \text{número de restricciones lineales}$$
$$k = \text{número de parámetros (incluido el intercepto) en la regresión no restringida}$$
$$n = \text{número de observaciones}$$

La matriz correspondiente a (8.6.9) es entonces

$$F = \frac{(\hat{\mathbf{u}}_R'\hat{\mathbf{u}}_R - \hat{\mathbf{u}}_{NR}'\hat{\mathbf{u}}_{NR})/m}{(\hat{\mathbf{u}}_{NR}'\hat{\mathbf{u}}_{NR})/(n-k)} \tag{C.8.1}$$

que sigue la distribución $F$ con $(m, n-k)$gl. Como es usual, si el valor $F$ calculado a partir de (C.8.1) excede al valor $F$ crítico, rechazamos la regresión restringida; de lo contrario, no la rechazamos.

## C.9 Predicción mediante regresión múltiple: formulación matricial

En la sección 8.8 analizamos, mediante notación escalar, la forma en que la regresión múltiple estimada predice (1) la media y (2) los valores individuales de $Y$, dados los valores de las regresoras $X$. En esta sección veremos cómo expresar estas predicciones en forma matricial. También presentamos las fórmulas para estimar las varianzas y los errores estándar de los valores pronosticados; en el capítulo 8 vimos que estas fórmulas se trabajan mejor mediante notación matricial, pues las expresiones escalares o algebraicas de estas fórmulas llegan a ser inmanejables.

### Predicción media

Sea

$$\mathbf{X_0} = \begin{bmatrix} 1 \\ X_{02} \\ X_{03} \\ \vdots \\ X_{0k} \end{bmatrix} \tag{C.9.1}$$

el vector de valores de las variables $X$ para las cuales se desea predecir $\hat{Y}_0$, la predicción media de $Y$.

Ahora la regresión múltiple estimada en forma escalar es

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + u_i \quad \text{(C.9.2)}$$

la cual en notación matricial se escribe de manera compacta como:

$$\hat{Y}_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}} \quad \text{(C.9.3)}$$

donde $\mathbf{x}_i' = [1 \; X_{2i} \; X_{3i} \; \cdots \; X_{ki}]$ y

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

La ecuación (C.9.2) o (C.9.3) es, por supuesto, la predicción media de $Y_i$ que corresponde a una $\mathbf{x}_i'$ dada.

Si $\mathbf{x}_i'$ es igual a la de (C.9.1), (C.9.3) se convierte en

$$(\hat{Y}_i \mid \mathbf{x}_0') = \mathbf{x}_0' \hat{\boldsymbol{\beta}} \quad \text{(C.9.4)}$$

donde, por supuesto, los valores de $\mathbf{x}_0$ están especificados. Observe que (C.9.4) da una predicción insesgada de $E(Y_i \mid \mathbf{x}_0')$, pues $E(x_0'\hat{\boldsymbol{\beta}}) = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$. (¿Por qué?)

## Varianza de la predicción media

La fórmula para estimar la varianza de $(\hat{Y}_0 \mid \mathbf{x}_0')$ es la siguiente:[7]

$$\text{var}\,(\hat{Y}_0 \mid \mathbf{x}_0') = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \quad \text{(C.9.5)}$$

donde $\sigma^2$ es la varianza de $u_i$, $\mathbf{x}_0'$ son los valores dados de las variables $X$ para los cuales se desea predecir, y $(\mathbf{X}'\mathbf{X})$ es la matriz dada en (C.3.9). En la práctica reemplazamos $\sigma^2$ por su estimador insesgado $\hat{\sigma}^2$.

En la siguiente sección ilustramos la predicción media y su varianza.

## Predicción individual

Como sabemos de los capítulos 5 y 8, la predicción individual de $Y(= Y_0)$ también está dada por (C.9.3), o en forma más específica por (C.9.4). La diferencia entre las predicciones de la media y la individual consiste en sus varianzas.

## Varianza de la predicción individual

La fórmula para la varianza de una predicción individual es la siguiente:[8]

$$\text{var}\,(Y_0 \mid \mathbf{x}_0) = \sigma^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0] \quad \text{(C.9.6)}$$

donde $\text{var}(Y_0 \mid \mathbf{x}_0)$ representa $E[Y_0 - \hat{Y}_0 \mid X]^2$. En la práctica, sustituimos $\sigma^2$ por su estimador insesgado $\hat{\sigma}^2$. En la siguiente sección ilustraremos esta fórmula.

[7] Para la derivación, véase J. Johnston, *Econometric Methods*, McGraw-Hill, 3a. ed., Nueva York, 1984, pp. 195-196.

[8] *Ibid.*

## C.10 Resumen del método matricial: un ejemplo ilustrativo

Considere los datos de la tabla C.4, los cuales pertenecen al gasto de consumo personal *per cápita* (GCPP) y al ingreso personal disponible *per cápita* (IPDP), así como al tiempo o la variable de tendencia. Al incluir esta última en el modelo, tratamos de averiguar la relación del GCPP con el IPDP neto de la variable de tendencia (que puede representar una multitud de otros factores, como tecnología, cambio en gustos, etcétera).

Para fines empíricos, por consiguiente, el modelo de regresión es

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{C.10.1}$$

donde $Y$ = gasto de consumo *per cápita*, $X_2$ = ingreso disponible *per cápita* y $X_3$ = tiempo. La información requerida para efectuar la regresión (C.10.1) se proporciona en la tabla C.4.

En notación matricial, el problema puede mostrarse de la siguiente manera:

$$\begin{bmatrix} 1\,673 \\ 1\,688 \\ 1\,666 \\ 1\,735 \\ 1\,749 \\ 1\,756 \\ 1\,815 \\ 1\,867 \\ 1\,948 \\ 2\,048 \\ 2\,128 \\ 2\,165 \\ 2\,257 \\ 2\,316 \\ 2\,324 \end{bmatrix} = \begin{bmatrix} 1 & 1\,839 & 1 \\ 1 & 1\,844 & 2 \\ 1 & 1\,831 & 3 \\ 1 & 1\,881 & 4 \\ 1 & 1\,883 & 5 \\ 1 & 1\,910 & 6 \\ 1 & 1\,969 & 7 \\ 1 & 2\,016 & 8 \\ 1 & 2\,126 & 9 \\ 1 & 2\,239 & 10 \\ 1 & 2\,336 & 11 \\ 1 & 2\,404 & 12 \\ 1 & 2\,487 & 13 \\ 1 & 2\,535 & 14 \\ 1 & 2\,595 & 15 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_9 \\ \hat{u}_{10} \\ \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{14} \\ \hat{u}_{15} \end{bmatrix} \tag{C.10.2}$$

$$\underset{15\times 1}{\mathbf{y}} = \underset{15\times 3}{\mathbf{X}} \quad \underset{3\times 1}{\hat{\boldsymbol{\beta}}} + \underset{15\times 1}{\hat{\mathbf{u}}}$$

**TABLA C.4**
**Gasto de consumo personal *per cápita* (GCPP) e ingreso personal disponible *per cápita* (IPDP) en Estados Unidos, 1956-1970, en dólares de 1958**

Fuente: *Economic Report of the President*, enero de 1972, tabla B-16.

| GCPP, $Y$ | IPDP, $X_2$ | Tiempo, $X_3$ | GCPP, $Y$ | IPDP, $X_2$ | Tiempo, $X_3$ |
|---|---|---|---|---|---|
| 1 673 | 1 839 | 1 ( = 1956) | 1 948 | 2 126 | 9 |
| 1 688 | 1 844 | 2 | 2 048 | 2 239 | 10 |
| 1 666 | 1 831 | 3 | 2 128 | 2 336 | 11 |
| 1 735 | 1 881 | 4 | 2 165 | 2 404 | 12 |
| 1 749 | 1 883 | 5 | 2 257 | 2 487 | 13 |
| 1 756 | 1 910 | 6 | 2 316 | 2 535 | 14 |
| 1 815 | 1 969 | 7 | 2 324 | 2 595 | 15 ( = 1970) |
| 1 867 | 2 016 | 8 | | | |

De los datos anteriores obtenemos los siguientes resultados:

$$\bar{Y} = 1\,942.333 \qquad \bar{X}_2 = 2\,126.333 \qquad \bar{X}_3 = 8.0$$

$$\sum (Y_i - \bar{Y})^2 = 830\ 121.333$$

$$\sum (X_{2i} - \bar{X}_2)^2 = 1\ 103\ 111.333 \qquad \sum (X_{3i} - \bar{X}_3)^2 = 280.0$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ 1 & X_{23} & X_{33} \\ \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{2i}X_{3i} & \sum X_{3i}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 15 & 31\ 895 & 120 \\ 31\ 895 & 68\ 922.513 & 272\ 144 \\ 120 & 272\ 144 & 1\ 240 \end{bmatrix} \tag{C.10.3}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 29\ 135 \\ 62\ 905\ 821 \\ 247\ 934 \end{bmatrix} \tag{C.10.4}$$

Mediante las reglas de inversión de matrices del **apéndice B** se ve que

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 37.232491 & -0.0225082 & 1.336707 \\ -0.0225082 & 0.0000137 & -0.0008319 \\ 1.336707 & -0.0008319 & 0.054034 \end{bmatrix} \tag{C.10.5}$$

Por consiguiente,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 300.28625 \\ 0.74198 \\ 8.04356 \end{bmatrix} \tag{C.10.6}$$

La suma de cuadrados residual ahora se calcula así:

$$\begin{aligned} \sum \hat{u}_i^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= 57\ 420\ 003 - [300.28625 \quad 0.74198 \quad 8.04356] \begin{bmatrix} 29\ 135 \\ 62\ 905\ 821 \\ 247\ 934 \end{bmatrix} \\ &= 1\ 976.85574 \end{aligned} \tag{C.10.7}$$

de donde obtenemos

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{12} = 164.73797 \tag{C.10.8}$$

Por tanto, la matriz de varianza-covarianza para $\hat{\boldsymbol{\beta}}$ se presenta como

$$\text{var-cov}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 6\ 133.650 & -3.70794 & 220.20634 \\ -3.70794 & 0.00226 & -0.13705 \\ 220.20634 & -0.13705 & 8.90155 \end{bmatrix} \tag{C.10.9}$$

Los elementos diagonales de esta matriz dan las varianzas de $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$, respectivamente, y sus raíces cuadradas positivas dan los errores estándar correspondientes.

De los datos anteriores se verifica fácilmente que

$$\text{SCE: } \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = 828\ 144.47786 \tag{C.10.10}$$

$$\text{SCT: } \mathbf{y}'\mathbf{y} - n\bar{Y}^2 = 830\ 121.333 \tag{C.10.11}$$

Por consiguiente,

$$\begin{aligned} R^2 &= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \\ &= \frac{828\ 144.47786}{830\ 121.333} \\ &= 0.99761 \end{aligned} \tag{C.10.12}$$

Aplicamos (7.8.4) y se ve que **el coeficiente de determinación ajustado** es

$$\bar{R}^2 = 0.99722 \tag{C.10.13}$$

Reunimos los resultados obtenidos hasta ahora y tenemos

$$\begin{aligned} \hat{Y}_i &= 300.28625 \quad + \quad 0.74198X_{2i} \quad + 8.04356X_{3i} \\ &\quad\ (78.31763) \qquad (0.04753) \qquad (2.98354) \\ t &= \quad (3.83421) \qquad (15.60956) \qquad (2.69598) \\ &\quad R^2 = 0.99761 \qquad \bar{R}^2 = 0.99722 \qquad \text{gl} = 12 \end{aligned} \tag{C.10.14}$$

La interpretación de (C.10.14) es: si tanto $X_2$ como $X_3$ reciben un valor fijo de cero, el valor promedio del gasto de consumo personal *per cápita* se estima alrededor de \$300. Como es usual, debemos tomar con cautela esta interpretación mecánica del intercepto. El coeficiente de regresión parcial de 0.74198 significa que, al mantener todas las otras variables constantes, un incremento en el ingreso *per cápita* de un dólar, por ejemplo, es acompañado por un incremento en el gasto de consumo personal *per cápita* medio de alrededor de 74 centavos de dólar. En resumen, se estima que la propensión marginal a consumir sea de alrededor de 0.74 o 74%. En forma similar, si mantenemos constantes todas las otras variables, el gasto de consumo personal *per cápita* medio aumentó a una tasa de alrededor de \$8 por año durante el periodo del estudio, 1956-1970. El valor $R^2$ de 0.9976 muestra que las dos variables explicativas representaron más de 99% de la variación en el gasto de consumo *per cápita* en Estados Unidos de 1956 a 1970. Aunque $\bar{R}^2$ se reduce un poco, continúa muy elevada.

De regreso a la significancia estadística de los coeficientes estimados, a partir de (C.10.14) observamos que cada coeficiente estimado es estadísticamente significativo *individualmente*, en un nivel de significancia de 5%, por ejemplo: las razones entre los coeficientes estimados y sus errores estándar (es decir, las razones *t*) son 3.83421, 15.61077 y 2.69598, respectivamente. Mediante una prueba *t* de dos colas en el nivel de significancia de 5% observamos que el valor *t* crítico para 12 gl es 2.179. Cada valor *t* calculado excede este valor crítico. Por tanto, en lo individual, podemos rechazar la hipótesis nula de que el verdadero valor poblacional del coeficiente relevante es cero.

Como ya vimos, no es posible aplicar la prueba *t* usual para verificar la hipótesis de que $\beta_2 = \beta_3 = 0$ simultáneamente, porque el procedimiento de prueba *t* supone tomar una muestra independiente cada vez que se aplica una prueba *t*. Si utilizamos la misma muestra para probar la hipótesis sobre $\beta_2$ y $\beta_3$ al mismo tiempo, es probable que los estimadores $\hat{\beta}_2$ y $\hat{\beta}_3$ estén corre-

**TABLA C.5**
**Tabla ANOVA para los datos de la tabla C.4**

| Origen de la variación | SC | gl | SCM |
|---|---|---|---|
| Debido a $X_2$, $X_3$ | 828 144.47786 | 2 | 414 072.3893 |
| Debido a los residuos | 1 976.85574 | 12 | 164.73797 |
| Total | 830 121.33360 | 14 | |

lacionados, lo que viola el supuesto en el cual se basa el procedimiento de la prueba $t$.[9] En realidad, la matriz de varianza-covarianza de $\hat{\boldsymbol{\beta}}$ de (C.10.9) muestra que los estimadores $\hat{\beta}_2$ y $\hat{\beta}_3$ están correlacionados negativamente (la covarianza entre los dos es $-0.13705$). Por tanto, no podemos utilizar la prueba $t$ para probar la hipótesis nula de que $\beta_2 = \beta_3 = 0$.

Sin embargo, recuerde que una hipótesis nula como $\beta_2 = \beta_3 = 0$, simultáneamente, se prueba con la técnica del análisis de varianza y la tradicional prueba $F$, que presentamos en el capítulo 8. Para este problema, la tabla del análisis de varianza es la tabla C.5. Según los supuestos usuales, obtenemos

$$F = \frac{414\ 072.3893}{164.73797} = 2\ 513.52 \qquad \text{(C.10.15)}$$

cuya distribución es igual a la distribución $F$ con 2 y 12 gl. El valor $F$ calculado es, obviamente, muy significativo; rechazamos la hipótesis nula de que $\beta_2 = \beta_3 = 0$, es decir, el gasto de consumo personal *per cápita* no está relacionado linealmente con el ingreso disponible *per cápita* ni con la tendencia.

En la sección C.9 estudiamos el mecanismo de pronóstico, de la media e individual. Suponga que para 1971 la cifra del IPDP es $2 610 y deseamos pronosticar el GCPP correspondiente a esta cifra. Entonces, la proyección media y la individual del GCPP para 1971 es la misma y está dada por

$$\begin{aligned}(\text{PPCE}_{1971} \mid \text{IPDP}_{1971}, X_3 = 16) &= \mathbf{x}'_{1971}\hat{\boldsymbol{\beta}} \\ &= [1 \quad 2\ 610 \quad 16]\begin{bmatrix} 300.28625 \\ 0.74198 \\ 8.04356 \end{bmatrix} \\ &= 2\ 365.55\end{aligned} \qquad \text{(C.10.16)}$$

donde empleamos la ecuación (C.9.3).

Como sabemos por la sección C.9, las varianzas de $\hat{Y}_{1971}$ y $Y_{1971}$ son diferentes y son las siguientes:

$$\begin{aligned}\text{var}\,(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) &= \hat{\sigma}^2[\mathbf{x}'_{1971}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{1971}] \\ &= 164.73797[1 \quad 2\ 610 \quad 16](\mathbf{X}'\mathbf{X})^{-1}\begin{bmatrix} 1 \\ 2\ 610 \\ 16 \end{bmatrix}\end{aligned} \qquad \text{(C.10.17)}$$

donde $(\mathbf{X}'\mathbf{X})^{-1}$ es como se muestra en (C.10.5). Al sustituir esto en (C.10.17), el lector debe verificar que

$$\text{var}\,(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) = 48.6426 \qquad \text{(C.10.18)}$$

[9] Para mayor detalle, véase la sección 8.4.

y por consiguiente,

$$\text{ee}\,(\hat{Y}_{1971}\,|\,\mathbf{x}'_{1971}) = 6.9744$$

Dejamos al lector verificar, mediante (C.9.6), que

$$\text{var}\,(Y_{1971}\,|\,\mathbf{x}'_{1971}) = 213.3806 \tag{C.10.19}$$

y

$$\text{ee}\,(Y_{1971}\,|\,\mathbf{x}'_{1971}) = 14.6076$$

*Nota:* $\text{var}\,(Y_{1971}\,|\,\mathbf{x}'_{1971}) = E[Y_{1971} - \hat{Y}_{1971}\,|\,\mathbf{x}'_{1971}]^2$.

En la sección C.5 presentamos la matriz de correlación ***R***. Para los datos, la matriz de correlación es la siguiente:

$$R = \begin{matrix} & Y & X_2 & X_3 \\ Y & 1 & 0.9980 & 0.9743 \\ X_2 & 0.9980 & 1 & 0.9664 \\ X_3 & 0.9743 & 0.9664 & 1 \end{matrix} \tag{C.10.20}$$

Observe que en (C.10.20) colocamos al margen de la matriz de correlación las variables del modelo, de forma que se identifiquen fácilmente las variables incluidas en el cálculo del coeficiente de correlación. Así, el coeficiente 0.9980 en el primer renglón de la matriz (C.10.20) es el coeficiente de correlación entre $Y$ y $X_2$ (es decir, $r_{12}$). A partir de las correlaciones de orden cero de la matriz de correlación (C.10.20) se derivan con facilidad los coeficientes de correlación de primer orden (véase el ejercicio C.7).

## C.11 Mínimos cuadrados generalizados (MCG)

En diversas ocasiones mencionamos que los MCO son un caso especial de los MCG. Para apreciar esto, regresemos a la ecuación (C.2.2). Con el objeto de tomar en cuenta las varianzas heteroscedásticas [los elementos de la diagonal principal de (C.2.2)] y las autocorrelaciones en los términos de error [los elementos que no están en la diagonal principal de (C.2.2)], suponemos que

$$\text{E}(\mathbf{uu}') = \sigma^2\mathbf{V} \tag{C.11.1}$$

donde $\mathbf{V}$ es una matriz de $n \times n$ conocida.

En consecuencia, si el modelo es:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

donde $E(\mathbf{u}) = \mathbf{0}$ y var-cov$(\mathbf{u}) = \sigma^2\mathbf{V}$. En caso de que no se conozca $\sigma^2$, lo cual suele suceder, $\mathbf{V}$ representa la estructura supuesta de las varianzas y de las covarianzas entre los errores aleatorios $u_t$.

Conforme a la condición impuesta de la varianza-covarianza de los términos de error, se demuestra que:

$$\boldsymbol{\beta}^{\text{mcg}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{C.11.2}$$

$\boldsymbol{\beta}^{\text{mcg}}$ se conoce como **estimador de mínimos cuadrados generalizados (MCG)** de $\boldsymbol{\beta}$.

También se demuestra que

$$\text{var-cov}\,(\boldsymbol{\beta}^{\text{mcg}}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{C.11.3}$$

Se puede probar que $\boldsymbol{\beta}^{\text{mcg}}$ es el mejor estimador lineal insesgado de $\boldsymbol{\beta}$.

Si se supone que la varianza de cada término de error es la misma constante $\sigma^2$ y que los términos de error están mutuamente no correlacionados, la matriz **V** se reduce a la matriz identidad, como se ve en (C.2.3). Si los términos de error no están mutuamente correlacionados pero tienen varianzas distintas (es decir, heteroscedásticas), la matriz **V** será diagonal, con varianzas diferentes a lo largo de la diagonal principal. Desde luego, si hay heteroscedasticidad y autocorrelación, la matriz **V** tendrá entradas en la diagonal principal y fuera de ella.

El verdadero problema en la práctica consiste en que desconocemos $\sigma^2$, así como las varianzas y covarianzas verdaderas (es decir, la estructura de la matriz **V**). Como una solución, recurrimos al método de **mínimos cuadrados generalizados estimados, o factibles (MCGE)**. En dicho método calculamos primero el modelo mediante MCO y pasamos por alto los problemas de la heteroscedasticidad y/o la autocorrelación. Obtenemos los residuos a partir de este modelo y de la matriz de varianza-covarianza (estimada) del término de error mediante la sustitución de las entradas de la expresión que está justamente antes de (C.2.2) por los $u$ estimados, a saber: $\hat{u}$. Se puede demostrar que los estimadores de MCGE son estimadores consistentes de los MCG. Simbólicamente,

$$\boldsymbol{\beta}^{\text{mcge}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}) \tag{C.11.4}$$

$$\text{var-cov}\,(\boldsymbol{\beta}^{\text{mcge}}) = \sigma^2(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \tag{C.11.5}$$

donde $\hat{\mathbf{V}}$ es una estimación de **V**.

## C.12 Resumen y conclusiones

El propósito principal de este apéndice fue introducir el enfoque matricial al modelo clásico de regresión lineal. Aunque presentamos muy pocos conceptos nuevos del análisis de regresión, la notación matricial proporciona un método compacto para abordar los modelos de regresión lineal con cualquier número de variables.

Al concluir este apéndice observe que si las variables $X$ y $Y$ se miden en forma de desviaciones, es decir, como desviaciones de sus medias muestrales, hay pocos cambios en las fórmulas presentadas antes. Estos cambios se enumeran en la tabla C.6.[10] Como muestra dicha tabla, en

**TABLA C.6**
**Modelo de regresión con *k* variables en unidades originales y en forma de desviación***

| Unidades originales | | Formas de desviación | |
|---|---|---|---|
| $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ | (C.3.2) | $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$<br>La columna de 1 en la matriz **X** se elimina (¿Por qué?) | |
| $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ | (C.3.11) | Igual | |
| $\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ | (C.3.13) | Igual | |
| $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ | (C.3.18) | Igual | |
| $\sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2$ | (C.3.16) | $\sum y_i^2 = \mathbf{y}'\mathbf{y}$ | (C.12.1) |
| $\text{SCE} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$ | (C.3.17) | $\text{SCE} = \boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ | (C.12.2) |
| $R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2}$ | (C.4.2) | $R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}$ | (C.12.3) |

* Observe que, a pesar de que en ambos casos los símbolos para las matrices y los vectores son iguales, en la forma de desviación se supone que los elementos de las matrices y de los vectores son desviaciones y no datos simples. Observe además que en la forma de desviación $\hat{\boldsymbol{\beta}}$ es de orden $k - 1$ y la var-cov($\hat{\boldsymbol{\beta}}$) es de orden $(k-1)(k-1)$.

[10] En estos días de computadoras de alta velocidad quizá no sea necesaria la forma de desviación; pero simplifica las fórmulas y, por consiguiente, los cálculos si trabaja con calculadora de escritorio y maneja cifras grandes.

la forma de desviaciones, se elimina de la SCT y de la SCE la corrección para la media $n\bar{Y}^2$. (¿Por qué?) Esta pérdida genera un cambio en la fórmula para $R^2$. Por lo demás, la mayoría de las fórmulas desarrolladas en las unidades originales de medición se mantiene para la forma de desviación.

## EJERCICIOS

C.1. Para el ejemplo ilustrativo analizado en la sección C.10, la $\mathbf{X}'\mathbf{X}$ y la $\mathbf{X}'\mathbf{y}$, con la información en forma de desviación, son las siguientes:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1\ 103\ 111.333 & 16\ 984 \\ 16\ 984 & 280 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 955\ 099.333 \\ 14\ 854.000 \end{bmatrix}$$

*a*) Estime $\beta_2$ y $\beta_3$.

*b*) ¿Cómo estimaría $\beta_1$?

*c*) Estime la varianza de $\hat{\beta}_2$ y $\hat{\beta}_3$ y sus covarianzas.

*d*) Obtenga $R^2$ y $\bar{R}^2$.

*e*) Al comparar los resultados con los de la sección C.10, ¿qué ventajas encuentra en la forma de desviación?

C.2. Consulte el ejercicio 22.23. Con los datos proporcionados en dicho ejercicio, defina la matriz $(\mathbf{X}'\mathbf{X})$ y el vector $\mathbf{X}'\mathbf{y}$; asimismo, calcule el vector parámetro $\boldsymbol{\beta}$ y su matriz de varianza-covarianza. También obtenga $R^2$. ¿Cómo demostraría la hipótesis de que las elasticidades de M1 respecto del PIB y la tasa de interés $R$ son numéricamente iguales?

C.3. *Prueba sobre la igualdad de dos coeficientes de regresión.* Suponga que tenemos el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

y desea probar la hipótesis de que $\beta_2 = \beta_3$. Si supone que los $u_i$ están normalmente distribuidos, se demuestra que

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\operatorname{var}(\hat{\beta}_2) + \operatorname{var}(\hat{\beta}_3) - 2\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3)}}$$

sigue la distribución $t$ con $n - 3$ gl (véase la sección 8.5). (En general, para el caso de $k$ variables, los gl son $n - k$.) Por consiguiente, con la prueba $t$ anterior se prueba la hipótesis nula $\beta_2 = \beta_3$.

Con la prueba $t$ anterior, verifique la hipótesis de que los verdaderos valores de $\beta_2$ y $\beta_3$ en la regresión (C.10.14) son idénticos.

*Sugerencia:* utilice la matriz var-cov de $\boldsymbol{\beta}$ dada en (C.10.9).

C.4. *Forma de expresar correlaciones de orden superior en términos de correlaciones de orden inferior.* Los coeficientes de correlación de orden $p$ se expresan en términos de los coeficientes de correlación de orden $p - 1$ mediante la siguiente **fórmula de reducción**:

$$r_{12.345\ldots p} = \frac{r_{12.345\ldots(p-1)} - \left[r_{1p.345\ldots(p-1)}r_{2p.345\ldots(p-1)}\right]}{\sqrt{\left[1 - r^2_{1p.345\ldots(p-1)}\right]}\sqrt{\left[1 - r^2_{2p.345\ldots(p-1)}\right]}}$$

Así,

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

igual a la obtenida en el capítulo 7.

Con la siguiente matriz de correlación:

$$\mathbf{R} = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 1 & 0.44 & -0.34 & -0.31 & -0.14 \\ & 1 & 0.25 & -0.19 & -0.35 \\ & & 1 & 0.44 & 0.33 \\ & & & 1 & 0.85 \\ & & & & 1 \end{bmatrix} \end{array}$$

Encuentre lo siguiente:

*a*) $r_{12.345}$ *b*) $r_{12.34}$ *c*) $r_{12.3}$

*d*) $r_{13.245}$ *e*) $r_{13.24}$ *f*) $r_{13.2}$

C.5. *Forma de expresar coeficientes de regresión de orden superior en términos de coeficientes de regresión de orden inferior.* Un coeficiente de regresión de orden $p$ se expresa en términos de un coeficiente de regresión de orden $p - 1$ mediante la siguiente fórmula de reducción:

$$\hat{\beta}_{12.345\ldots p} = \frac{\hat{\beta}_{12.345\ldots(p-1)} - \left[\hat{\beta}_{1p.345\ldots(p-1)}\hat{\beta}_{p2.345\ldots(p-1)}\right]}{1 - \hat{\beta}_{2p.345\ldots(p-1)}\hat{\beta}_{p2.345\ldots(p-1)}}$$

Así,

$$\hat{\beta}_{12.3} = \frac{\hat{\beta}_{12} - \hat{\beta}_{13}\hat{\beta}_{32}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}}$$

donde $\beta_{12.3}$ es el coeficiente de la pendiente en la regresión de $y$ sobre $X_2$ si $X_3$ se mantiene constante. En forma similar, $\beta_{12.34}$ es el coeficiente de la pendiente en la regresión de $Y$ sobre $X_2$ si se mantienen constantes $X_3$ y $X_4$, y así sucesivamente.

Con la fórmula anterior, encuentre expresiones para los siguientes coeficientes de regresión en términos de coeficientes de regresión de orden menor: $\hat{\beta}_{12.3456}$, $\hat{\beta}_{12.345}$ y $\hat{\beta}_{12.34}$.

C.6. Establezca la siguiente identidad:

$$\hat{\beta}_{12.3}\hat{\beta}_{23.1}\hat{\beta}_{31.2} = r_{12.3}r_{23.1}r_{31.2}$$

C.7. Para la matriz de correlación $\boldsymbol{R}$ de (C.10.20), encuentre todos los coeficientes de correlación parcial de primer orden.

C.8. Al estudiar la variación en las tasas de criminalidad en algunas ciudades grandes de Estados Unidos, Ogburn obtuvo la siguiente información:*

| | |
|---|---|
| $\bar{Y} = 19.9$ | $S_1 = 7.9$ |
| $\bar{X}_2 = 49.2$ | $S_2 = 1.3$ |
| $\bar{X}_3 = 10.2$ | $S_3 = 4.6$ |
| $\bar{X}_4 = 481.4$ | $S_4 = 74.4$ |
| $\bar{X}_5 = 41.6$ | $S_5 = 10.8$ |

$$\boldsymbol{R} = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 1 & 0.44 & -0.34 & -0.31 & -0.14 \\ & 1 & 0.25 & -0.19 & -0.35 \\ & & 1 & 0.44 & 0.33 \\ & & & 1 & 0.85 \\ & & & & 1 \end{bmatrix} \end{array}$$

* W.F. Ogburn, "Factors in the Variation of Crime among Cities", *Journal of American Statistical Association*, vol. 30, 1935, p. 12.

donde $Y =$ tasa de criminalidad, número de delitos conocidos por cada mil personas
$X_2 =$ porcentaje de habitantes masculinos
$X_3 =$ porcentaje del total de habitantes masculinos nacidos en el exterior
$X_4 =$ número de niños menores de 5 años por cada mil mujeres casadas entre los 15 y los 44 años de edad
$X_5 =$ pertenencia a alguna iglesia, número de miembros de la iglesia de 13 años de edad o mayores por cada 100 personas de la población total de 13 años de edad y superior; de $S_1$ a $S_5$ son las desviaciones estándar muestrales de las variables $Y$ hasta $X_5$, y $\boldsymbol{R}$ es la matriz de correlación.

*a*) Trate $Y$ como variable dependiente, obtenga la regresión de $Y$ sobre las cuatro variables $X$ e interprete la regresión estimada.

*b*) Obtenga $r_{12.3}$, $r_{14.35}$ y $r_{15.34}$.

*c*) Obtenga $R^2$ y pruebe la hipótesis de que todos los coeficientes de pendiente parciales son simultáneamente iguales a cero.

C.9. En la siguiente tabla se proporcionan datos sobre la producción y costo total de producción de un bien en el corto plazo (véase el ejemplo 7.4).

| Producción | Costo total, $ |
|---|---|
| 1 | 193 |
| 2 | 226 |
| 3 | 240 |
| 4 | 244 |
| 5 | 257 |
| 6 | 260 |
| 7 | 274 |
| 8 | 297 |
| 9 | 350 |
| 10 | 420 |

Para probar si los datos anteriores sugieren las curvas de costo promedio y de costo marginal en forma de U que por lo general se encuentra en el corto plazo, se puede utilizar el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$$

donde $Y =$ costo total y $X =$ producción. Las variables explicativas adicionales $X_i^2$ y $X_i^3$ son derivadas de $X$.

*a*) Exprese los datos en la forma de desviación y obtenga $(\mathbf{X'X})$,$(\mathbf{X'y})$ y $(\mathbf{X'X})^{-1}$.

*b*) Estime $\beta_2$, $\beta_3$ y $\beta_4$.

*c*) Estime la matriz var-cov de $\hat{\boldsymbol{\beta}}$.

*d*) Estime $\beta_1$. Interprete $\hat{\beta}_1$ en el contexto del problema.

*e*) Obtenga $R^2$ y $\bar{R}^2$.

*f*) *A priori*, ¿cuáles son los signos de $\beta_2$, $\beta_3$ y $\beta_4$? ¿Por qué?

*g*) De la función de costo total anterior, obtenga expresiones para las funciones de costo marginal y promedio.

*h*) Ajuste las funciones de costo promedio y marginal a los datos y comente sobre el ajuste.

*i*) Si $\beta_3 = \beta_4 = 0$, ¿cuál es la naturaleza de la función de costo marginal? ¿Cómo probaría la hipótesis de que $\beta_3 = \beta_4 = 0$?

*j*) ¿Cómo derivaría las funciones de costo variable total y de costo variable promedio a partir de la información dada?

**TABLA C.7**
**Experiencia de la participación de la población urbana de bajos recursos en la fuerza laboral: área del censo, ciudad de Nueva York, 1970**

Fuente: Áreas del censo: Nueva York, Bureau of the Census, U.S. Department of Commerce, 1970.

| Núm. de área | % en la fuerza laboral $Y$* | Ingreso familiar medio, $X_2$† | Tamaño familiar medio, $X_3$ | Tasa de desempleo, $X_4$‡ |
|---|---|---|---|---|
| 137 | 64.3 | 1 998 | 2.95 | 4.4 |
| 139 | 45.4 | 1 114 | 3.40 | 3.4 |
| 141 | 26.6 | 1 942 | 3.72 | 1.1 |
| 142 | 87.5 | 1 998 | 4.43 | 3.1 |
| 143 | 71.3 | 2 026 | 3.82 | 7.7 |
| 145 | 82.4 | 1 853 | 3.90 | 5.0 |
| 147 | 26.3 | 1 666 | 3.32 | 6.2 |
| 149 | 61.6 | 1 434 | 3.80 | 5.4 |
| 151 | 52.9 | 1 513 | 3.49 | 12.2 |
| 153 | 64.7 | 2 008 | 3.85 | 4.8 |
| 155 | 64.9 | 1 704 | 4.69 | 2.9 |
| 157 | 70.5 | 1 525 | 3.89 | 4.8 |
| 159 | 87.2 | 1 842 | 3.53 | 3.9 |
| 161 | 81.2 | 1 735 | 4.96 | 7.2 |
| 163 | 67.9 | 1 639 | 3.68 | 3.6 |

*$Y$ = jefes de familia menores de 65 años.
†$X_2$ = dólares.
‡$X_4$ = porcentaje de la fuerza laboral civil desempleada.

C.10. Con el fin de estudiar la participación de las familias urbanas de bajos recursos (con ingresos menores a $3 943 en 1969) en la fuerza laboral, se obtuvieron los siguientes datos del Censo de Población de 1970.

*a*) Con el modelo de regresión $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$ obtenga las estimaciones de los coeficientes de regresión e interprete sus resultados.

*b*) *A priori*, ¿cuáles son los signos esperados de los coeficientes de regresión en el modelo anterior y por qué?

*c*) ¿Cómo probaría la hipótesis de que la tasa global de desempleo no afecta la participación en la fuerza laboral de la población urbana de bajos recursos en el área del censo de la tabla anterior?

*d*) ¿Debe eliminarse alguna variable del modelo anterior? ¿Por qué?

*e*) ¿Qué otras variables deben incluirse en el modelo?

C.11. En una aplicación de la función de producción Cobb-Douglas se obtuvieron los siguientes resultados:

$$\widehat{\ln Y}_i = 2.3542 + \underset{(0.3022)}{0.9576} \ln X_{2i} + \underset{(0.3571)}{0.8242} \ln X_{3i}$$

$$R^2 = 0.8432 \qquad \text{gl} = 12$$

donde $Y$ = producción, $X_2$ = insumo trabajo y $X_3$ = insumo capital, y donde las cifras en paréntesis son los errores estándar estimados.

*a*) Como mencionamos en el capítulo 7, los coeficientes de los insumos trabajo y capital en la ecuación anterior dan las elasticidades de la producción respecto del trabajo y el capital. Pruebe la hipótesis de que estas elasticidades son *individualmente* iguales a la unidad.

*b*) Pruebe la hipótesis de que las elasticidades trabajo y capital son iguales, suponiendo *i*) que la covarianza entre los coeficientes estimados del trabajo y del capital es cero y *ii*) que es −0.0972.

*c*) ¿Cómo probaría la significancia global de la ecuación de regresión anterior?

*C.12. Exprese la función de verosimilitud para el modelo de regresión con $k$ variables en notación matricial y muestre que $\tilde{\boldsymbol{\beta}}$, el vector de estimadores de máxima verosimilitud, es idéntico a $\hat{\boldsymbol{\beta}}$ el vector de estimadores de MCO del modelo de regresión con $k$ variables.

C.13. *Regresión mediante variables estandarizadas.* Considere las siguientes funciones de regresión muestral (FRM):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \qquad (1)$$

$$Y_i^* = b_1 + b_2 X_{2i}^* + b_3 X_{3i}^* + \hat{u}_i^* \qquad (2)$$

donde

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_Y}$$
$$X_{2i}^* = \frac{X_{2i} - \bar{X}_2}{s_2}$$
$$X_{3i}^* = \frac{X_{3i} - \bar{X}_3}{s_3}$$

donde las letras $s$ denotan las desviaciones estándar muestrales. Como vimos en el capítulo 6, sección 6.3, las variables con asterisco se conocen como *variables estandarizadas*. Tienen medias cero y desviaciones estándar unitarias ($= 1$). Exprese todas las variables en forma de desviaciones y demuestre lo siguiente para el modelo (2):

*a*) $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & r_{23} \\ r_{23} & 1 \end{bmatrix} n$

*b*) $\mathbf{X}'\mathbf{y} = \begin{bmatrix} r_{12} \\ r_{13} \end{bmatrix} n$

*c*) $\mathbf{X}'\mathbf{X}^{-1} = \dfrac{1}{n\left(1 - r_{23}^2\right)} \begin{bmatrix} 1 & -r_{23} \\ -r_{23} & 1 \end{bmatrix}$

*d*) $\hat{\boldsymbol{\beta}} = \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \dfrac{1}{1 - r_{23}^2} \begin{bmatrix} r_{12} - r_{23} r_{13} \\ r_{13} - r_{23} r_{12} \end{bmatrix}$

*e*) $b_1 = 0$

Establezca también la relación entre las $b$ y las $\hat{\beta}$.

(Observe que en las relaciones anteriores, $n$ denota el tamaño de la muestra; $r_{12}$, $r_{13}$ y $r_{23}$ denotan las correlaciones entre $Y$ y $X_2$, entre $Y$ y $X_3$ y entre $X_2$ y $X_3$, respectivamente.)

C.14. Verifique las ecuaciones (C.10.18) y (C.10.19).

*C.15. *Mínimos cuadrados restringidos.* Suponga que

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (1)$$

la cual deseamos estimar sujeta a un conjunto de limitaciones o restricciones de igualdad:

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \qquad (2)$$

donde $\mathbf{R}$ es una matriz *conocida* de orden $qxk$ ($q \leq k$) y $\mathbf{r}$ es un vector *conocido* de $q$ elementos. Para ilustrar, suponga que nuestro modelo es

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \qquad (3)$$

* Opcional.

y suponga que deseamos estimar este modelo sujeto a estas restricciones:

$$\begin{aligned} \beta_2 - \beta_3 &= 0 \\ \beta_4 + \beta_5 &= 1 \end{aligned} \tag{4}$$

Con algunas de las técnicas estudiadas en el capítulo 8 podemos incorporar estas restricciones (por ejemplo, $\beta_2 = \beta_3$ y $\beta_4 = 1 - \beta_5$, y eliminar así $\beta_2$ y $\beta_4$ del modelo) y probar la validez de estas restricciones mediante la prueba $F$ allí tratada. Pero una forma más directa de estimar (3) incorporando las restricciones (4) directamente en el procedimiento de estimación es expresar primero las restricciones en la forma de la ecuación (2), que en el presente caso se convierte en

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5}$$

Al permitir que $\boldsymbol{\beta}^*$ denote el estimador de mínimos cuadrados restringidos, se demuestra que $\boldsymbol{\beta}^*$ se estima mediante la siguiente fórmula:*

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}) \tag{6}$$

donde $\hat{\boldsymbol{\beta}}$ es el estimador usual (no restringido) calculado mediante la fórmula usual $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

*a*) ¿Cuál es el vector $\beta$ en (3)?

*b*) Con este vector $\beta$, verifique que la matriz **R** y el vector **r** dados en (5) incorporan en realidad las restricciones especificadas en (4).

*c*) Escriba la matriz **R** y el vector **r** en los siguientes casos:

*i*) $\beta_2 = \beta_3 = \beta_4 = 2$

*ii*) $\beta_2 = \beta_3$ y $\beta_4 = \beta_5$

*iii*) $\beta_2 - 3\beta_3 = 5\beta_4$

*iv*) $\beta_2 + 3\beta_3 = 0$

*d*) ¿Cuándo será $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$?

# Apéndice CA

## CA.1 Derivación de *k* ecuaciones normales o simultáneas

Al diferenciar

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

parcialmente respecto de $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$, obtenemos

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-1)$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-X_{2i})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_k} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{ki} - \cdots - \hat{\beta}_k X_{ki})(-X_{ki})$$

Igualamos a cero las derivadas parciales anteriores, reordenamos los términos y obtenemos las $k$ ecuaciones normales dadas en (C.3.8).

* Véase J. Johnston, *op. cit.*, p. 205.

## CA.2 Derivación matricial de las ecuaciones normales

A partir de (C.3.7) obtenemos

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Con las reglas de diferenciación matricial del **apéndice B, sección B.6**, obtenemos

$$\frac{\partial(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Igualamos a cero la ecuación anterior y obtenemos

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

de donde $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ siempre que exista la inversa.

## CA.3 Matriz de varianza-covarianza de $\hat{\boldsymbol{\beta}}$

De (C.3.11) obtenemos

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Sustituimos $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ en la expresión anterior y obtenemos

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (1) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

Por consiguiente,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (2)$$

Por definición,

$$\begin{aligned}\text{var-cov}\,(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\ &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} \qquad (3) \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]\end{aligned}$$

donde en el último paso aprovechamos que $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Como las $X$ son no estocásticas, al tomar el valor esperado de (3) obtenemos:

$$\begin{aligned}\text{var-cov}\,(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

que es el resultado de (C.3.13). Observe que al derivar el resultado anterior empleamos el supuesto de que $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$.

## CA.4 Propiedad MELI de los estimadores de MCO

De (C.3.11) tenemos que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (1)$$

Como $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ es una matriz de números fijos, $\hat{\boldsymbol{\beta}}$ es una función lineal de $Y$. Por tanto, por definición, es un estimador lineal.

Recuerde que la FRP es

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

Sustituimos esto en (1) y obtenemos

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \tag{3}$$

$$= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \tag{4}$$

porque $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$.

Tomamos el valor esperado de (4) y da

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u})$$

$$= \boldsymbol{\beta} \tag{5}$$

porque $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ (¿por qué?) y $E(\mathbf{u}) = \mathbf{0}$ según los supuestos, lo cual indica que $\hat{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$.

Sea $\hat{\boldsymbol{\beta}}^*$ cualquier otro estimador lineal de $\boldsymbol{\beta}$, el cual se escribe como

$$\boldsymbol{\beta}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]\mathbf{y} \tag{6}$$

donde $\mathbf{C}$ es una matriz de constantes.

Al sustituir $\mathbf{y}$ de (2) en (6), obtenemos

$$\begin{aligned}\hat{\boldsymbol{\beta}}^* &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}](\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= \boldsymbol{\beta} + \mathbf{C}\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u}\end{aligned} \tag{7}$$

Ahora, si $\hat{\boldsymbol{\beta}}^*$ es un estimador insesgado de $\boldsymbol{\beta}$, debemos tener

$$\mathbf{C}\mathbf{X} = 0 \qquad \text{(¿Por qué?)} \tag{8}$$

Con (8), (7) se escribe así:

$$\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u} \tag{9}$$

Por definición, la matriz de var-cov($\hat{\boldsymbol{\beta}}^*$) es

$$E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})' = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u}]' \tag{10}$$

Empleamos las propiedades de inversión y trasposición de matrices y después de simplificación algebraica para obtener

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}^*) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{C}'$$

$$= \text{var-cov}\,(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{C}\mathbf{C}' \tag{11}$$

lo cual indica que la matriz de varianza-covarianza del estimador lineal e insesgado alterno $\hat{\boldsymbol{\beta}}^*$ es igual a la matriz de varianza-covarianza del estimador MCO, $\hat{\boldsymbol{\beta}}$ más $\sigma^2$ veces $\mathbf{C}\mathbf{C}'$, que es una matriz semidefinida* positiva. Por tanto, las varianzas de un elemento dado de $\hat{\boldsymbol{\beta}}^*$ deben ser necesariamente iguales o mayores al elemento correspondiente de $\hat{\boldsymbol{\beta}}$, lo cual demuestra que $\hat{\boldsymbol{\beta}}$ es MELI. Por supuesto, si $\mathbf{C}$ es una matriz nula, es decir, $\mathbf{C} = \mathbf{0}$, entonces $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$, lo que equivale a decir que si encontramos un estimador MELI, éste debe ser el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$.

* Consulte las referencias del **apéndice B**.

# Apéndice D

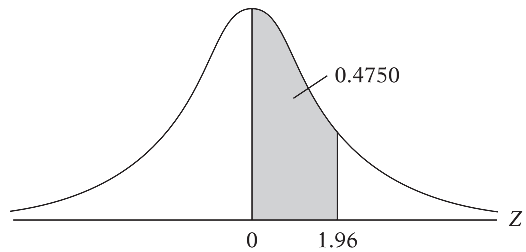
# Tablas estadísticas

## TABLA D.1
**Áreas debajo de la distribución normal estandarizada**

**Ejemplo**

$\Pr(0 \leq Z \leq 1.96) = 0.4750$

$\Pr(Z \geq 1.96) = 0.5 - 0.4750 = 0.025$




| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| 0.7 | .2580 | .2611 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4454 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |

*Nota:* Esta tabla da el área en la cola del lado derecho de la distribución (es decir, $Z \geq 0$). Pero como la distribución normal es simétrica alrededor de $Z = 0$, el área en la cola del lado izquierdo es la misma que el área en la cola correspondiente del lado derecho. Por ejemplo, $P(-1.96 \leq Z \leq 0) = 0.4750$. Por consiguiente, $P(-1.96 \leq Z \leq 1.96) = 2(0.4750) = 0.95$.

## TABLA D.2
**Puntos porcentuales de la distribución *t***

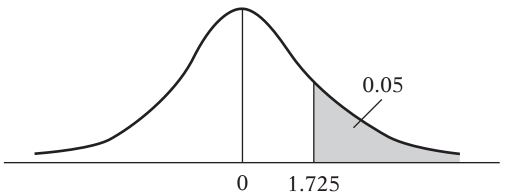
Fuente: De E. S. Pearson y H. O. Hartley (eds.), *Biometrika Tables for Statisticians*, vol. 1, 3a. ed., tabla 12, Cambridge University Press, Nueva York, 1966. Se reproduce con autorización de los editores y los fideicomisarios de *Biometrika*.

**Ejemplo**

$\Pr(t > 2.086) = 0.025$

$\Pr(t > 1.725) = 0.05$ para gl = 20

$\Pr(|t| > 1.725) = 0.10$




| Pr / gl | 0.25<br>0.50 | 0.10<br>0.20 | 0.05<br>0.10 | 0.025<br>0.05 | 0.01<br>0.02 | 0.005<br>0.010 | 0.001<br>0.002 |
|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.31 |
| 2 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 |
| 3 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.214 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 |
| 5 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 |
| 6 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 |
| 7 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 |
| 8 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 |
| 9 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 |
| 10 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 |
| 11 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 |
| 12 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 |
| 13 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 |
| 14 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 |
| 15 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 |
| 16 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 |
| 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 |
| 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 |
| 19 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 |
| 20 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 |
| 21 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 |
| 22 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 |
| 23 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 |
| 24 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 |
| 25 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 |
| 26 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 |
| 27 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 |
| 28 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 |
| 29 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 |
| 30 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 |
| 40 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 |
| 60 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 |
| 120 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 |
| $\infty$ | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |

*Nota:* La probabilidad más baja que aparece en el encabezado de cada columna es el área en una cola; la probabilidad más alta es el área en ambas colas.

## TABLA D.3 Puntos porcentuales superiores de la distribución *F*

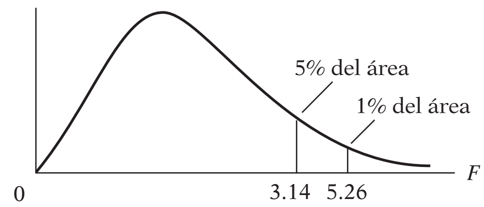
**Ejemplo**

$\Pr(F > 1.59) = 0.25$

$\Pr(F > 2.42) = 0.10$ **para** gl $N_1 = 10$

$\Pr(F > 3.14) = 0.05$ y $N_2 = 9$

$\Pr(F > 5.26) = 0.01$




| gl para el denominador $N_2$ | gl para el numerador $N_1$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pr | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | .25 | 5.83 | 7.50 | 8.20 | 8.58 | 8.82 | 8.98 | 9.10 | 9.19 | 9.26 | 9.32 | 9.36 | 9.41 |
| | .10 | 39.9 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 | 58.9 | 59.4 | 59.9 | 60.2 | 60.5 | 60.7 |
| | .05 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 243 | 244 |
| 2 | .25 | 2.57 | 3.00 | 3.15 | 3.23 | 3.28 | 3.31 | 3.34 | 3.35 | 3.37 | 3.38 | 3.39 | 3.39 |
| | .10 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 | 9.39 | 9.40 | 9.41 |
| | .05 | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 |
| | .01 | 98.5 | 99.0 | 99.2 | 99.2 | 99.3 | 99.3 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 | 99.4 |
| 3 | .25 | 2.02 | 2.28 | 2.36 | 2.39 | 2.41 | 2.42 | 2.43 | 2.44 | 2.44 | 2.44 | 2.45 | 2.45 |
| | .10 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 | 5.23 | 5.22 | 5.22 |
| | .05 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.76 | 8.74 |
| | .01 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 | 27.5 | 27.3 | 27.2 | 27.1 | 27.1 |
| 4 | .25 | 1.81 | 2.00 | 2.05 | 2.06 | 2.07 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 |
| | .10 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 | 3.92 | 3.91 | 3.90 |
| | .05 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.94 | 5.91 |
| | .01 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 | 14.8 | 14.7 | 14.5 | 14.4 | 14.4 |
| 5 | .25 | 1.69 | 1.85 | 1.88 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 |
| | .10 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 | 3.30 | 3.28 | 3.27 |
| | .05 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.71 | 4.68 |
| | .01 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 | 10.3 | 10.2 | 10.1 | 9.96 | 9.89 |
| 6 | .25 | 1.62 | 1.76 | 1.78 | 1.79 | 1.79 | 1.78 | 1.78 | 1.78 | 1.77 | 1.77 | 1.77 | 1.77 |
| | .10 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 | 2.94 | 2.92 | 2.90 |
| | .05 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.03 | 4.00 |
| | .01 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.79 | 7.72 |
| 7 | .25 | 1.57 | 1.70 | 1.72 | 1.72 | 1.71 | 1.71 | 1.70 | 1.70 | 1.69 | 1.69 | 1.69 | 1.68 |
| | .10 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 | 2.70 | 2.68 | 2.67 |
| | .05 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.60 | 3.57 |
| | .01 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.54 | 6.47 |
| 8 | .25 | 1.54 | 1.66 | 1.67 | 1.66 | 1.66 | 1.65 | 1.64 | 1.64 | 1.63 | 1.63 | 1.63 | 1.62 |
| | .10 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 | 2.54 | 2.52 | 2.50 |
| | .05 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.31 | 3.28 |
| | .01 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.73 | 5.67 |
| 9 | .25 | 1.51 | 1.62 | 1.63 | 1.63 | 1.62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.58 |
| | .10 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 | 2.42 | 2.40 | 2.38 |
| | .05 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.10 | 3.07 |
| | .01 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.18 | 5.11 |

(*continúa*)

*Fuente:* De E.S. Pearson y H.O. Hartley (eds.), *Biometrika Tables for Statisticians*, vol. 1, 3a. ed., tabla 18, Cambridge University Press, Nueva York, 1966. Se reproduce con autorización de los editores y los fideicomisarios de *Biometrika*.

(*continuación*)

| gl para el numerador $N_1$ | | | | | | | | | | | | | gl para el denominador $N_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **15** | **20** | **24** | **30** | **40** | **50** | **60** | **100** | **120** | **200** | **500** | **∞** | **Pr** | |
| 9.49 | 9.58 | 9.63 | 9.67 | 9.71 | 9.74 | 9.76 | 9.78 | 9.80 | 9.82 | 9.84 | 9.85 | .25 | |
| 61.2 | 61.7 | 62.0 | 62.3 | 62.5 | 62.7 | 62.8 | 63.0 | 63.1 | 63.2 | 63.3 | 63.3 | .10 | 1 |
| 246 | 248 | 249 | 250 | 251 | 252 | 252 | 253 | 253 | 254 | 254 | 254 | .05 | |
| 3.41 | 3.43 | 3.43 | 3.44 | 3.45 | 3.45 | 3.46 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 | .25 | |
| 9.42 | 9.44 | 9.45 | 9.46 | 9.47 | 9.47 | 9.47 | 9.48 | 9.48 | 9.49 | 9.49 | 9.49 | .10 | 2 |
| 19.4 | 19.4 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 | .05 | |
| 99.4 | 99.4 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | 99.5 | .01 | |
| 2.46 | 2.46 | 2.46 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | 2.47 | .25 | |
| 5.20 | 5.18 | 5.18 | 5.17 | 5.16 | 5.15 | 5.15 | 5.14 | 5.14 | 5.14 | 5.14 | 5.13 | .10 | 3 |
| 8.70 | 8.66 | 8.64 | 8.62 | 8.59 | 8.58 | 8.57 | 8.55 | 8.55 | 8.54 | 8.53 | 8.53 | .05 | |
| 26.9 | 26.7 | 26.6 | 26.5 | 26.4 | 26.4 | 26.3 | 26.2 | 26.2 | 26.2 | 26.1 | 26.1 | .01 | |
| 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | .25 | |
| 3.87 | 3.84 | 3.83 | 3.82 | 3.80 | 3.80 | 3.79 | 3.78 | 3.78 | 3.77 | 3.76 | 3.76 | .10 | 4 |
| 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.70 | 5.69 | 5.66 | 5.66 | 5.65 | 5.64 | 5.63 | .05 | |
| 14.2 | 14.0 | 13.9 | 13.8 | 13.7 | 13.7 | 13.7 | 13.6 | 13.6 | 13.5 | 13.5 | 13.5 | .01 | |
| 1.89 | 1.88 | 1.88 | 1.88 | 1.88 | 1.88 | 1.87 | 1.87 | 1.87 | 1.87 | 1.87 | 1.87 | .25 | |
| 3.24 | 3.21 | 3.19 | 3.17 | 3.16 | 3.15 | 3.14 | 3.13 | 3.12 | 3.12 | 3.11 | 3.10 | .10 | 5 |
| 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.44 | 4.43 | 4.41 | 4.40 | 4.39 | 4.37 | 4.36 | .05 | |
| 9.72 | 9.55 | 9.47 | 9.38 | 9.29 | 9.24 | 9.20 | 9.13 | 9.11 | 9.08 | 9.04 | 9.02 | .01 | |
| 1.76 | 1.76 | 1.75 | 1.75 | 1.75 | 1.75 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 | .25 | |
| 2.87 | 2.84 | 2.82 | 2.80 | 2.78 | 2.77 | 2.76 | 2.75 | 2.74 | 2.73 | 2.73 | 2.72 | .10 | 6 |
| 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.75 | 3.74 | 3.71 | 3.70 | 3.69 | 3.68 | 3.67 | .05 | |
| 7.56 | 7.40 | 7.31 | 7.23 | 7.14 | 7.09 | 7.06 | 6.99 | 6.97 | 6.93 | 6.90 | 6.88 | .01 | |
| 1.68 | 1.67 | 1.67 | 1.66 | 1.66 | 1.66 | 1.65 | 1.65 | 1.65 | 1.65 | 1.65 | 1.65 | .25 | |
| 2.63 | 2.59 | 2.58 | 2.56 | 2.54 | 2.52 | 2.51 | 2.50 | 2.49 | 2.48 | 2.48 | 2.47 | .10 | 7 |
| 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.32 | 3.30 | 3.27 | 3.27 | 3.25 | 3.24 | 3.23 | .05 | |
| 6.31 | 6.16 | 6.07 | 5.99 | 5.91 | 5.86 | 5.82 | 5.75 | 5.74 | 5.70 | 5.67 | 5.65 | .01 | |
| 1.62 | 1.61 | 1.60 | 1.60 | 1.59 | 1.59 | 1.59 | 1.58 | 1.58 | 1.58 | 1.58 | 1.58 | .25 | |
| 2.46 | 2.42 | 2.40 | 2.38 | 2.36 | 2.35 | 2.34 | 2.32 | 2.32 | 2.31 | 2.30 | 2.29 | .10 | 8 |
| 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 2.02 | 3.01 | 2.97 | 2.97 | 2.95 | 2.94 | 2.93 | .05 | |
| 5.52 | 5.36 | 5.28 | 5.20 | 5.12 | 5.07 | 5.03 | 4.96 | 4.95 | 4.91 | 4.88 | 4.86 | .01 | |
| 1.57 | 1.56 | 1.56 | 1.55 | 1.55 | 1.54 | 1.54 | 1.53 | 1.53 | 1.53 | 1.53 | 1.53 | .25 | |
| 2.34 | 2.30 | 2.28 | 2.25 | 2.23 | 2.22 | 2.21 | 2.19 | 2.18 | 2.17 | 2.17 | 2.16 | .10 | 9 |
| 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2.80 | 2.79 | 2.76 | 2.75 | 2.73 | 2.72 | 2.71 | .05 | |
| 4.96 | 4.81 | 4.73 | 4.65 | 4.57 | 4.52 | 4.48 | 4.42 | 4.40 | 4.36 | 4.33 | 4.31 | .01 | |

(*continúa*)

**TABLA D.3 Puntos porcentuales superiores de la distribución *F* (*continuación*)**

| gl para el denominador $N_2$ | gl para el numerador $N_1$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pr | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 10 | .25 | 1.49 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.57 | 1.56 | 1.56 | 1.55 | 1.55 | 1.54 |
| | .10 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 | 2.30 | 2.28 |
| | .05 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.94 | 2.91 |
| | .01 | 10.0 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.77 | 4.71 |
| 11 | .25 | 1.47 | 1.58 | 1.58 | 1.57 | 1.56 | 1.55 | 1.54 | 1.53 | 1.53 | 1.52 | 1.52 | 1.51 |
| | .10 | 3.23 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 | 2.23 | 2.21 |
| | .05 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.82 | 2.79 |
| | .01 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.46 | 4.40 |
| 12 | .25 | 1.46 | 1.56 | 1.56 | 1.55 | 1.54 | 1.53 | 1.52 | 1.51 | 1.51 | 1.50 | 1.50 | 1.49 |
| | .10 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.21 | 2.19 | 2.17 | 2.15 |
| | .05 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.72 | 2.69 |
| | .01 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.22 | 4.16 |
| 13 | .25 | 1.45 | 1.55 | 1.55 | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 |
| | .10 | 3.14 | 2.76 | 2.56 | 2.43 | 2.35 | 2.28 | 2.23 | 2.20 | 2.16 | 2.14 | 2.12 | 2.10 |
| | .05 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 | 2.63 | 2.60 |
| | .01 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 | 4.02 | 3.96 |
| 14 | .25 | 1.44 | 1.53 | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.48 | 1.47 | 1.46 | 1.46 | 1.45 |
| | .10 | 3.10 | 2.73 | 2.52 | 2.39 | 2.31 | 2.24 | 2.19 | 2.15 | 2.12 | 2.10 | 2.08 | 2.05 |
| | .05 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 | 2.57 | 2.53 |
| | .01 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.86 | 3.80 |
| 15 | .25 | 1.43 | 1.52 | 1.52 | 1.51 | 1.49 | 1.48 | 1.47 | 1.46 | 1.46 | 1.45 | 1.44 | 1.44 |
| | .10 | 3.07 | 2.70 | 2.49 | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.09 | 2.06 | 2.04 | 2.02 |
| | .05 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.51 | 2.48 |
| | .01 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.73 | 3.67 |
| 16 | .25 | 1.42 | 1.51 | 1.51 | 1.50 | 1.48 | 1.47 | 1.46 | 1.45 | 1.44 | 1.44 | 1.44 | 1.43 |
| | .10 | 3.05 | 2.67 | 2.46 | 2.33 | 2.24 | 2.18 | 2.13 | 2.09 | 2.06 | 2.03 | 2.01 | 1.99 |
| | .05 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.46 | 2.42 |
| | .01 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.62 | 3.55 |
| 17 | .25 | 1.42 | 1.51 | 1.50 | 1.49 | 1.47 | 1.46 | 1.45 | 1.44 | 1.43 | 1.43 | 1.42 | 1.41 |
| | .10 | 3.03 | 2.64 | 2.44 | 2.31 | 2.22 | 2.15 | 2.10 | 2.06 | 2.03 | 2.00 | 1.98 | 1.96 |
| | .05 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.41 | 2.38 |
| | .01 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.52 | 3.46 |
| 18 | .25 | 1.41 | 1.50 | 1.49 | 1.48 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 | 1.41 | 1.40 |
| | .10 | 3.01 | 2.62 | 2.42 | 2.29 | 2.20 | 2.13 | 2.08 | 2.04 | 2.00 | 1.98 | 1.96 | 1.93 |
| | .05 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.37 | 2.34 |
| | .01 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.43 | 3.37 |
| 19 | .25 | 1.41 | 1.49 | 1.49 | 1.47 | 1.46 | 1.44 | 1.43 | 1.42 | 1.41 | 1.41 | 1.40 | 1.40 |
| | .10 | 2.99 | 2.61 | 2.40 | 2.27 | 2.18 | 2.11 | 2.06 | 2.02 | 1.98 | 1.96 | 1.94 | 1.91 |
| | .05 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.34 | 2.31 |
| | .01 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.36 | 3.30 |
| 20 | .25 | 1.40 | 1.49 | 1.48 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.39 |
| | .10 | 2.97 | 2.59 | 2.38 | 2.25 | 2.16 | 2.09 | 2.04 | 2.00 | 1.96 | 1.94 | 1.92 | 1.89 |
| | .05 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.31 | 2.28 |
| | .01 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.29 | 3.23 |

(*continúa*)

(*continuación*)

| gl para el numerador $N_1$ | | | | | | | | | | | | | gl para el denominador $N_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 20 | 24 | 30 | 40 | 50 | 60 | 100 | 120 | 200 | 500 | ∞ | Pr | |
| 1.53 | 1.52 | 1.52 | 1.51 | 1.51 | 1.50 | 1.50 | 1.49 | 1.49 | 1.49 | 1.48 | 1.48 | .25 | 10 |
| 2.24 | 2.20 | 2.18 | 2.16 | 2.13 | 2.12 | 2.11 | 2.09 | 2.08 | 2.07 | 2.06 | 2.06 | .10 | |
| 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.64 | 2.62 | 2.59 | 2.58 | 2.56 | 2.55 | 2.54 | .05 | |
| 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.12 | 4.08 | 4.01 | 4.00 | 3.96 | 3.93 | 3.91 | .01 | |
| 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 | 1.47 | 1.46 | 1.46 | 1.46 | 1.45 | 1.45 | .25 | 11 |
| 2.17 | 2.12 | 2.10 | 2.08 | 2.05 | 2.04 | 2.03 | 2.00 | 2.00 | 1.99 | 1.98 | 1.97 | .10 | |
| 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.51 | 2.49 | 2.46 | 2.45 | 2.43 | 2.42 | 2.40 | .05 | |
| 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.81 | 3.78 | 3.71 | 3.69 | 3.66 | 3.62 | 3.60 | .01 | |
| 1.48 | 1.47 | 1.46 | 1.45 | 1.45 | 1.44 | 1.44 | 1.43 | 1.43 | 1.43 | 1.42 | 1.42 | .25 | 12 |
| 2.10 | 2.06 | 2.04 | 2.01 | 1.99 | 1.97 | 1.96 | 1.94 | 1.93 | 1.92 | 1.91 | 1.90 | .10 | |
| 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.40 | 2.38 | 2.35 | 2.34 | 2.32 | 2.31 | 2.30 | .05 | |
| 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.57 | 3.54 | 3.47 | 3.45 | 3.41 | 3.38 | 3.36 | .01 | |
| 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 | 1.42 | 1.41 | 1.41 | 1.40 | 1.40 | 1.40 | .25 | 13 |
| 2.05 | 2.01 | 1.98 | 1.96 | 1.93 | 1.92 | 1.90 | 1.88 | 1.88 | 1.86 | 1.85 | 1.85 | .10 | |
| 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.31 | 2.30 | 2.26 | 2.25 | 2.23 | 2.22 | 2.21 | .05 | |
| 3.82 | 3.66 | 3.59 | 3.51 | 3.43 | 3.38 | 3.34 | 3.27 | 3.25 | 3.22 | 3.19 | 3.17 | .01 | |
| 1.44 | 1.43 | 1.42 | 1.41 | 1.41 | 1.40 | 1.40 | 1.39 | 1.39 | 1.39 | 1.38 | 1.38 | .25 | 14 |
| 2.01 | 1.96 | 1.94 | 1.91 | 1.89 | 1.87 | 1.86 | 1.83 | 1.83 | 1.82 | 1.80 | 1.80 | .10 | |
| 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 2.24 | 2.22 | 2.19 | 2.18 | 2.16 | 2.14 | 2.13 | .05 | |
| 3.66 | 3.51 | 3.43 | 3.35 | 3.27 | 3.22 | 3.18 | 3.11 | 3.09 | 3.06 | 3.03 | 3.00 | .01 | |
| 1.43 | 1.41 | 1.41 | 1.40 | 1.39 | 1.39 | 1.38 | 1.38 | 1.37 | 1.37 | 1.36 | 1.36 | .25 | 15 |
| 1.97 | 1.92 | 1.90 | 1.87 | 1.85 | 1.83 | 1.82 | 1.79 | 1.79 | 1.77 | 1.76 | 1.76 | .10 | |
| 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.18 | 2.16 | 2.12 | 2.11 | 2.10 | 2.08 | 2.07 | .05 | |
| 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.08 | 3.05 | 2.98 | 2.96 | 2.92 | 2.89 | 2.87 | .01 | |
| 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.37 | 1.36 | 1.36 | 1.35 | 1.35 | 1.34 | 1.34 | .25 | 16 |
| 1.94 | 1.89 | 1.87 | 1.84 | 1.81 | 1.79 | 1.78 | 1.76 | 1.75 | 1.74 | 1.73 | 1.72 | .10 | |
| 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.12 | 2.11 | 2.07 | 2.06 | 2.04 | 2.02 | 2.01 | .05 | |
| 3.41 | 3.26 | 3.18 | 3.10 | 3.02 | 2.97 | 2.93 | 2.86 | 2.84 | 2.81 | 2.78 | 2.75 | .01 | |
| 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.35 | 1.34 | 1.34 | 1.34 | 1.33 | 1.33 | .25 | 17 |
| 1.91 | 1.86 | 1.84 | 1.81 | 1.78 | 1.76 | 1.75 | 1.73 | 1.72 | 1.71 | 1.69 | 1.69 | .10 | |
| 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.08 | 2.06 | 2.02 | 2.01 | 1.99 | 1.97 | 1.96 | .05 | |
| 3.31 | 3.16 | 3.08 | 3.00 | 2.92 | 2.87 | 2.83 | 2.76 | 2.75 | 2.71 | 2.68 | 2.65 | .01 | |
| 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.34 | 1.33 | 1.33 | 1.32 | 1.32 | 1.32 | .25 | 18 |
| 1.89 | 1.84 | 1.81 | 1.78 | 1.75 | 1.74 | 1.72 | 1.70 | 1.69 | 1.68 | 1.67 | 1.66 | .10 | |
| 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.04 | 2.02 | 1.98 | 1.97 | 1.95 | 1.93 | 1.92 | .05 | |
| 3.23 | 3.08 | 3.00 | 2.92 | 2.84 | 2.78 | 2.75 | 2.68 | 2.66 | 2.62 | 2.59 | 2.57 | .01 | |
| 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.33 | 1.32 | 1.32 | 1.31 | 1.31 | 1.30 | .25 | 19 |
| 1.86 | 1.81 | 1.79 | 1.76 | 1.73 | 1.71 | 1.70 | 1.67 | 1.67 | 1.65 | 1.64 | 1.63 | .10 | |
| 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 2.00 | 1.98 | 1.94 | 1.93 | 1.91 | 1.89 | 1.88 | .05 | |
| 3.15 | 3.00 | 2.92 | 2.84 | 2.76 | 2.71 | 2.67 | 2.60 | 2.58 | 2.55 | 2.51 | 2.49 | .01 | |
| 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.33 | 1.32 | 1.31 | 1.31 | 1.30 | 1.30 | 1.29 | .25 | 20 |
| 1.84 | 1.79 | 1.77 | 1.74 | 1.71 | 1.69 | 1.68 | 1.65 | 1.64 | 1.63 | 1.62 | 1.61 | .10 | |
| 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.97 | 1.95 | 1.91 | 1.90 | 1.88 | 1.86 | 1.84 | .05 | |
| 3.09 | 2.94 | 2.86 | 2.78 | 2.69 | 2.64 | 2.61 | 2.54 | 2.52 | 2.48 | 2.44 | 2.42 | .01 | |

(*continúa*)

**TABLA D.3 Puntos porcentuales superiores de la distribución *F* (*continuación*)**

| gl para el denominador $N_2$ | gl para el numerador $N_1$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pr | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 22 | .25 | 1.40 | 1.48 | 1.47 | 1.45 | 1.44 | 1.42 | 1.41 | 1.40 | 1.39 | 1.39 | 1.38 | 1.37 |
| | .10 | 2.95 | 2.56 | 2.35 | 2.22 | 2.13 | 2.06 | 2.01 | 1.97 | 1.93 | 1.90 | 1.88 | 1.86 |
| | .05 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.26 | 2.23 |
| | .01 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.18 | 3.12 |
| 24 | .25 | 1.39 | 1.47 | 1.46 | 1.44 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.38 | 1.37 | 1.36 |
| | .10 | 2.93 | 2.54 | 2.33 | 2.19 | 2.10 | 2.04 | 1.98 | 1.94 | 1.91 | 1.88 | 1.85 | 1.83 |
| | .05 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.21 | 2.18 |
| | .01 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.09 | 3.03 |
| 26 | .25 | 1.38 | 1.46 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.37 | 1.36 | 1.35 |
| | .10 | 2.91 | 2.52 | 2.31 | 2.17 | 2.08 | 2.01 | 1.96 | 1.92 | 1.88 | 1.86 | 1.84 | 1.81 |
| | .05 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.18 | 2.15 |
| | .01 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 3.02 | 2.96 |
| 28 | .25 | 1.38 | 1.46 | 1.45 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 |
| | .10 | 2.89 | 2.50 | 2.29 | 2.16 | 2.06 | 2.00 | 1.94 | 1.90 | 1.87 | 1.84 | 1.81 | 1.79 |
| | .05 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2.15 | 2.12 |
| | .01 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.96 | 2.90 |
| 30 | .25 | 1.38 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.35 | 1.34 |
| | .10 | 2.88 | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.93 | 1.88 | 1.85 | 1.82 | 1.79 | 1.77 |
| | .05 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.13 | 2.09 |
| | .01 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.91 | 2.84 |
| 40 | .25 | 1.36 | 1.44 | 1.42 | 1.40 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 |
| | .10 | 2.84 | 2.44 | 2.23 | 2.09 | 2.00 | 1.93 | 1.87 | 1.83 | 1.79 | 1.76 | 1.73 | 1.71 |
| | .05 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.04 | 2.00 |
| | .01 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.73 | 2.66 |
| 60 | .25 | 1.35 | 1.42 | 1.41 | 1.38 | 1.37 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.29 |
| | .10 | 2.79 | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.82 | 1.77 | 1.74 | 1.71 | 1.68 | 1.66 |
| | .05 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.95 | 1.92 |
| | .01 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.56 | 2.50 |
| 120 | .25 | 1.34 | 1.40 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 |
| | .10 | 2.75 | 2.35 | 2.13 | 1.99 | 1.90 | 1.82 | 1.77 | 1.72 | 1.68 | 1.65 | 1.62 | 1.60 |
| | .05 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.09 | 2.02 | 1.96 | 1.91 | 1.87 | 1.83 |
| | .01 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.40 | 2.34 |
| 200 | .25 | 1.33 | 1.39 | 1.38 | 1.36 | 1.34 | 1.32 | 1.31 | 1.29 | 1.28 | 1.27 | 1.26 | 1.25 |
| | .10 | 2.73 | 2.33 | 2.11 | 1.97 | 1.88 | 1.80 | 1.75 | 1.70 | 1.66 | 1.63 | 1.60 | 1.57 |
| | .05 | 3.89 | 3.04 | 2.65 | 2.42 | 2.26 | 2.14 | 2.06 | 1.98 | 1.93 | 1.88 | 1.84 | 1.80 |
| | .01 | 6.76 | 4.71 | 3.88 | 3.41 | 3.11 | 2.89 | 2.73 | 2.60 | 2.50 | 2.41 | 2.34 | 2.27 |
| $\infty$ | .25 | 1.32 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.29 | 1.28 | 1.27 | 1.25 | 1.24 | 1.24 |
| | .10 | 2.71 | 2.30 | 2.08 | 1.94 | 1.85 | 1.77 | 1.72 | 1.67 | 1.63 | 1.60 | 1.57 | 1.55 |
| | .05 | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.79 | 1.75 |
| | .01 | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.25 | 2.18 |

(*continúa*)

(*continuación*)

| gl para el numerador $N_1$ | | | | | | | | | | | | | gl para el denominador $N_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 20 | 24 | 30 | 40 | 50 | 60 | 100 | 120 | 200 | 500 | ∞ | Pr | |
| 1.36 | 1.34 | 1.33 | 1.32 | 1.31 | 1.31 | 1.30 | 1.30 | 1.30 | 1.29 | 1.29 | 1.28 | .25 | 22 |
| 1.81 | 1.76 | 1.73 | 1.70 | 1.67 | 1.65 | 1.64 | 1.61 | 1.60 | 1.59 | 1.58 | 1.57 | .10 | |
| 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.91 | 1.89 | 1.85 | 1.84 | 1.82 | 1.80 | 1.78 | .05 | |
| 2.98 | 2.83 | 2.75 | 2.67 | 2.58 | 2.53 | 2.50 | 2.42 | 2.40 | 2.36 | 2.33 | 2.31 | .01 | |
| 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.29 | 1.28 | 1.28 | 1.27 | 1.27 | 1.26 | .25 | 24 |
| 1.78 | 1.73 | 1.70 | 1.67 | 1.64 | 1.62 | 1.61 | 1.58 | 1.57 | 1.56 | 1.54 | 1.53 | .10 | |
| 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.86 | 1.84 | 1.80 | 1.79 | 1.77 | 1.75 | 1.73 | .05 | |
| 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.44 | 2.40 | 2.33 | 2.31 | 2.27 | 2.24 | 2.21 | .01 | |
| 1.34 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1.28 | 1.26 | 1.26 | 1.26 | 1.25 | 1.25 | .25 | 26 |
| 1.76 | 1.71 | 1.68 | 1.65 | 1.61 | 1.59 | 1.58 | 1.55 | 1.54 | 1.53 | 1.51 | 1.50 | .10 | |
| 2.07 | 1.99 | 1.95 | 1.90 | 1.85 | 1.82 | 1.80 | 1.76 | 1.75 | 1.73 | 1.71 | 1.69 | .05 | |
| 2.81 | 2.66 | 2.58 | 2.50 | 2.42 | 2.36 | 2.33 | 2.25 | 2.23 | 2.19 | 2.16 | 2.13 | .01 | |
| 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.27 | 1.26 | 1.25 | 1.25 | 1.24 | 1.24 | .25 | 28 |
| 1.74 | 1.69 | 1.66 | 1.63 | 1.59 | 1.57 | 1.56 | 1.53 | 1.52 | 1.50 | 1.49 | 1.48 | .10 | |
| 2.04 | 1.96 | 1.91 | 1.87 | 1.82 | 1.79 | 1.77 | 1.73 | 1.71 | 1.69 | 1.67 | 1.65 | .05 | |
| 2.75 | 2.60 | 2.52 | 2.44 | 2.35 | 2.30 | 2.26 | 2.19 | 2.17 | 2.13 | 2.09 | 2.06 | .01 | |
| 1.32 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.26 | 1.25 | 1.24 | 1.24 | 1.23 | 1.23 | .25 | 30 |
| 1.72 | 1.67 | 1.64 | 1.61 | 1.57 | 1.55 | 1.54 | 1.51 | 1.50 | 1.48 | 1.47 | 1.46 | .10 | |
| 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.76 | 1.74 | 1.70 | 1.68 | 1.66 | 1.64 | 1.62 | .05 | |
| 2.70 | 2.55 | 2.47 | 2.39 | 2.30 | 2.25 | 2.21 | 2.13 | 2.11 | 2.07 | 2.03 | 2.01 | .01 | |
| 1.30 | 1.28 | 1.26 | 1.25 | 1.24 | 1.23 | 1.22 | 1.21 | 1.21 | 1.20 | 1.19 | 1.19 | .25 | 40 |
| 1.66 | 1.61 | 1.57 | 1.54 | 1.51 | 1.48 | 1.47 | 1.43 | 1.42 | 1.41 | 1.39 | 1.38 | .10 | |
| 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.66 | 1.64 | 1.59 | 1.58 | 1.55 | 1.53 | 1.51 | .05 | |
| 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.06 | 2.02 | 1.94 | 1.92 | 1.87 | 1.83 | 1.80 | .01 | |
| 1.27 | 1.25 | 1.24 | 1.22 | 1.21 | 1.20 | 1.19 | 1.17 | 1.17 | 1.16 | 1.15 | 1.15 | .25 | 60 |
| 1.60 | 1.54 | 1.51 | 1.48 | 1.44 | 1.41 | 1.40 | 1.36 | 1.35 | 1.33 | 1.31 | 1.29 | .10 | |
| 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.56 | 1.53 | 1.48 | 1.47 | 1.44 | 1.41 | 1.39 | .05 | |
| 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.88 | 1.84 | 1.75 | 1.73 | 1.68 | 1.63 | 1.60 | .01 | |
| 1.24 | 1.22 | 1.21 | 1.19 | 1.18 | 1.17 | 1.16 | 1.14 | 1.13 | 1.12 | 1.11 | 1.10 | .25 | 120 |
| 1.55 | 1.48 | 1.45 | 1.41 | 1.37 | 1.34 | 1.32 | 1.27 | 1.26 | 1.24 | 1.21 | 1.19 | .10 | |
| 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.46 | 1.43 | 1.37 | 1.35 | 1.32 | 1.28 | 1.25 | .05 | |
| 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 1.70 | 1.66 | 1.56 | 1.53 | 1.48 | 1.42 | 1.38 | .01 | |
| 1.23 | 1.21 | 1.20 | 1.18 | 1.16 | 1.14 | 1.12 | 1.11 | 1.10 | 1.09 | 1.08 | 1.06 | .25 | 200 |
| 1.52 | 1.46 | 1.42 | 1.38 | 1.34 | 1.31 | 1.28 | 1.24 | 1.22 | 1.20 | 1.17 | 1.14 | .10 | |
| 1.72 | 1.62 | 1.57 | 1.52 | 1.46 | 1.41 | 1.39 | 1.32 | 1.29 | 1.26 | 1.22 | 1.19 | .05 | |
| 2.13 | 1.97 | 1.89 | 1.79 | 1.69 | 1.63 | 1.58 | 1.48 | 1.44 | 1.39 | 1.33 | 1.28 | .01 | |
| 1.22 | 1.19 | 1.18 | 1.16 | 1.14 | 1.13 | 1.12 | 1.09 | 1.08 | 1.07 | 1.04 | 1.00 | .25 | ∞ |
| 1.49 | 1.42 | 1.38 | 1.34 | 1.30 | 1.26 | 1.24 | 1.18 | 1.17 | 1.13 | 1.08 | 1.00 | .10 | |
| 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.35 | 1.32 | 1.24 | 1.22 | 1.17 | 1.11 | 1.00 | .05 | |
| 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.52 | 1.47 | 1.36 | 1.32 | 1.25 | 1.15 | 1.00 | .01 | |

## TABLA D.4

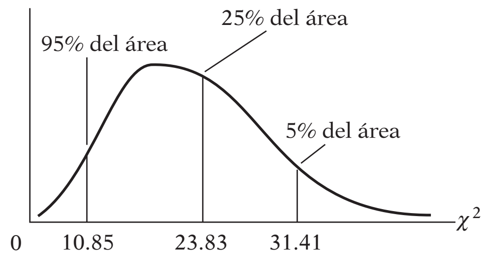
**Puntos porcentuales superiores de la distribución $\chi^2$**

**Ejemplo**

$\Pr(\chi^2 > 10.85) = 0.95$

$\Pr(\chi^2 > 23.83) = 0.25$ para gl = 20

$\Pr(\chi^2 > 31.41) = 0.05$




| Grados de libertad \ Pr | .995 | .990 | .975 | .950 | .900 |
|---|---|---|---|---|---|
| 1 | $392704 \times 10^{-10}$ | $157088 \times 10^{-9}$ | $982069 \times 10^{-9}$ | $393214 \times 10^{-8}$ | .0157908 |
| 2 | .0100251 | .0201007 | .0506356 | .102587 | .210720 |
| 3 | .0717212 | .114832 | .215795 | .351846 | .584375 |
| 4 | .206990 | .297110 | .484419 | .710721 | 1.063623 |
| 5 | .411740 | .554300 | .831211 | 1.145476 | 1.61031 |
| 6 | .675727 | .872085 | 1.237347 | 1.63539 | 2.20413 |
| 7 | .989265 | 1.239043 | 1.68987 | 2.16735 | 2.83311 |
| 8 | 1.344419 | 1.646482 | 2.17973 | 2.73264 | 3.48954 |
| 9 | 1.734926 | 2.087912 | 2.70039 | 3.32511 | 4.16816 |
| 10 | 2.15585 | 2.55821 | 3.24697 | 3.94030 | 4.86518 |
| 11 | 2.60321 | 3.05347 | 3.81575 | 4.57481 | 5.57779 |
| 12 | 3.07382 | 3.57056 | 4.40379 | 5.22603 | 6.30380 |
| 13 | 3.56503 | 4.10691 | 5.00874 | 5.89186 | 7.04150 |
| 14 | 4.07468 | 4.66043 | 5.62872 | 6.57063 | 7.78953 |
| 15 | 4.60094 | 5.22935 | 6.26214 | 7.26094 | 8.54675 |
| 16 | 5.14224 | 5.81221 | 6.90766 | 7.96164 | 9.31223 |
| 17 | 5.69724 | 6.40776 | 7.56418 | 8.67176 | 10.0852 |
| 18 | 6.26481 | 7.01491 | 8.23075 | 9.39046 | 10.8649 |
| 19 | 6.84398 | 7.63273 | 8.90655 | 10.1170 | 11.6509 |
| 20 | 7.43386 | 8.26040 | 9.59083 | 10.8508 | 12.4426 |
| 21 | 8.03366 | 8.89720 | 10.28293 | 11.5913 | 13.2396 |
| 22 | 8.64272 | 9.54249 | 10.9823 | 12.3380 | 14.0415 |
| 23 | 9.26042 | 10.19567 | 11.6885 | 13.0905 | 14.8479 |
| 24 | 9.88623 | 10.8564 | 12.4011 | 13.8484 | 15.6587 |
| 25 | 10.5197 | 11.5240 | 13.1197 | 14.6114 | 16.4734 |
| 26 | 11.1603 | 12.1981 | 13.8439 | 15.3791 | 17.2919 |
| 27 | 11.8076 | 12.8786 | 14.5733 | 16.1513 | 18.1138 |
| 28 | 12.4613 | 13.5648 | 15.3079 | 16.9279 | 18.9392 |
| 29 | 13.1211 | 14.2565 | 16.0471 | 17.7083 | 19.7677 |
| 30 | 13.7867 | 14.9535 | 16.7908 | 18.4926 | 20.5992 |
| 40 | 20.7065 | 22.1643 | 24.4331 | 26.5093 | 29.0505 |
| 50 | 27.9907 | 29.7067 | 32.3574 | 34.7642 | 37.6886 |
| 60 | 35.5346 | 37.4848 | 40.4817 | 43.1879 | 46.4589 |
| 70 | 43.2752 | 45.4418 | 48.7576 | 51.7393 | 55.3290 |
| 80 | 51.1720 | 53.5400 | 57.1532 | 60.3915 | 64.2778 |
| 90 | 59.1963 | 61.7541 | 65.6466 | 69.1260 | 73.2912 |
| 100* | 67.3276 | 70.0648 | 74.2219 | 77.9295 | 82.3581 |

(*continúa*)

* Para gl mayores que 100, la expresión $\sqrt{2\chi^2} - \sqrt{(2k-1)} = Z$ sigue la distribución normal estandarizada, donde $k$ representa los grados de libertad.

(*continuación*)

| .750 | .500 | .250 | .100 | .050 | .025 | .010 | .005 |
|---|---|---|---|---|---|---|---|
| .1015308 | .454937 | 1.32330 | 2.70554 | 3.84146 | 5.02389 | 6.63490 | 7.87944 |
| .575364 | 1.38629 | 2.77259 | 4.60517 | 5.99147 | 7.37776 | 9.21034 | 10.5966 |
| 1.212534 | 2.36597 | 4.10835 | 6.25139 | 7.81473 | 9.34840 | 11.3449 | 12.8381 |
| 1.92255 | 3.35670 | 5.38527 | 7.77944 | 9.48773 | 11.1433 | 13.2767 | 14.8602 |
| 2.67460 | 4.35146 | 6.62568 | 9.23635 | 11.0705 | 12.8325 | 15.0863 | 16.7496 |
| 3.45460 | 5.34812 | 7.84080 | 10.6446 | 12.5916 | 14.4494 | 16.8119 | 18.5476 |
| 4.25485 | 6.34581 | 9.03715 | 12.0170 | 14.0671 | 16.0128 | 18.4753 | 20.2777 |
| 5.07064 | 7.34412 | 10.2188 | 13.3616 | 15.5073 | 17.5346 | 20.0902 | 21.9550 |
| 5.89883 | 8.34283 | 11.3887 | 14.6837 | 16.9190 | 19.0228 | 21.6660 | 23.5893 |
| 6.73720 | 9.34182 | 12.5489 | 15.9871 | 18.3070 | 20.4831 | 23.2093 | 25.1882 |
| 7.58412 | 10.3410 | 13.7007 | 17.2750 | 19.6751 | 21.9200 | 24.7250 | 26.7569 |
| 8.43842 | 11.3403 | 14.8454 | 18.5494 | 21.0261 | 23.3367 | 26.2170 | 28.2995 |
| 9.29906 | 12.3398 | 15.9839 | 19.8119 | 22.3621 | 24.7356 | 27.6883 | 29.8194 |
| 10.1653 | 13.3393 | 17.1170 | 21.0642 | 23.6848 | 26.1190 | 29.1413 | 31.3193 |
| 11.0365 | 14.3389 | 18.2451 | 22.3072 | 24.9958 | 27.4884 | 30.5779 | 32.8013 |
| 11.9122 | 15.3385 | 19.3688 | 23.5418 | 26.2962 | 28.8454 | 31.9999 | 34.2672 |
| 12.7919 | 16.3381 | 20.4887 | 24.7690 | 27.5871 | 30.1910 | 33.4087 | 35.7185 |
| 13.6753 | 17.3379 | 21.6049 | 25.9894 | 28.8693 | 31.5264 | 34.8053 | 37.1564 |
| 14.5620 | 18.3376 | 22.7178 | 27.2036 | 30.1435 | 32.8523 | 36.1908 | 38.5822 |
| 15.4518 | 19.3374 | 23.8277 | 28.4120 | 31.4104 | 34.1696 | 37.5662 | 39.9968 |
| 16.3444 | 20.3372 | 24.9348 | 29.6151 | 32.6705 | 35.4789 | 38.9321 | 41.4010 |
| 17.2396 | 21.3370 | 26.0393 | 30.8133 | 33.9244 | 36.7807 | 40.2894 | 42.7956 |
| 18.1373 | 22.3369 | 27.1413 | 32.0069 | 35.1725 | 38.0757 | 41.6384 | 44.1813 |
| 19.0372 | 23.3367 | 28.2412 | 33.1963 | 36.4151 | 39.3641 | 42.9798 | 45.5585 |
| 19.9393 | 24.3366 | 29.3389 | 34.3816 | 37.6525 | 40.6465 | 44.3141 | 46.9278 |
| 20.8434 | 25.3364 | 30.4345 | 35.5631 | 38.8852 | 41.9232 | 45.6417 | 48.2899 |
| 21.7494 | 26.3363 | 31.5284 | 36.7412 | 40.1133 | 43.1944 | 46.9630 | 49.6449 |
| 22.6572 | 27.3363 | 32.6205 | 37.9159 | 41.3372 | 44.4607 | 48.2782 | 50.9933 |
| 23.5666 | 28.3362 | 33.7109 | 39.0875 | 42.5569 | 45.7222 | 49.5879 | 52.3356 |
| 24.4776 | 29.3360 | 34.7998 | 40.2560 | 43.7729 | 46.9792 | 50.8922 | 53.6720 |
| 33.6603 | 39.3354 | 45.6160 | 51.8050 | 55.7585 | 59.3417 | 63.6907 | 66.7659 |
| 42.9421 | 49.3349 | 56.3336 | 63.1671 | 67.5048 | 71.4202 | 76.1539 | 79.4900 |
| 52.2938 | 59.3347 | 66.9814 | 74.3970 | 79.0819 | 83.2976 | 88.3794 | 91.9517 |
| 61.6983 | 69.3344 | 77.5766 | 85.5271 | 90.5312 | 95.0231 | 100.425 | 104.215 |
| 71.1445 | 79.3343 | 88.1303 | 96.5782 | 101.879 | 106.629 | 112.329 | 116.321 |
| 80.6247 | 89.3342 | 98.6499 | 107.565 | 113.145 | 118.136 | 124.116 | 128.299 |
| 90.1332 | 99.3341 | 109.141 | 118.498 | 124.342 | 129.561 | 135.807 | 140.169 |

Fuente: Compendio de E.S. Pearson y H.O. Hartley (eds.), *Biometrika Tables for Statisticians*, vol. 1, 3a. ed., tabla 8, Cambridge University Press, Nueva York, 1966. Se reproduce con autorización de los editores y los fideicomisarios de *Biometrika*.

**TABLA D.5A** **Estadístico *d* de Durbin-Watson: puntos de significancia de $d_L$ y $d_U$ en el nivel de significancia de 0.05**

| | $k'=1$ | | $k'=2$ | | $k'=3$ | | $k'=4$ | | $k'=5$ | | $k'=6$ | | $k'=7$ | | $k'=8$ | | $k'=9$ | | $k'=10$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *n* | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 6 | 0.610 | 1.400 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 7 | 0.700 | 1.356 | 0.467 | 1.896 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 8 | 0.763 | 1.332 | 0.559 | 1.777 | 0.368 | 2.287 | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 9 | 0.824 | 1.320 | 0.629 | 1.699 | 0.455 | 2.128 | 0.296 | 2.588 | — | — | — | — | — | — | — | — | — | — | — | — |
| 10 | 0.879 | 1.320 | 0.697 | 1.641 | 0.525 | 2.016 | 0.376 | 2.414 | 0.243 | 2.822 | — | — | — | — | — | — | — | — | — | — |
| 11 | 0.927 | 1.324 | 0.658 | 1.604 | 0.595 | 1.928 | 0.444 | 2.283 | 0.316 | 2.645 | 0.203 | 3.005 | — | — | — | — | — | — | — | — |
| 12 | 0.971 | 1.331 | 0.812 | 1.579 | 0.658 | 1.864 | 0.512 | 2.177 | 0.379 | 2.506 | 0.268 | 2.832 | 0.171 | 3.149 | — | — | — | — | — | — |
| 13 | 1.010 | 1.340 | 0.861 | 1.562 | 0.715 | 1.816 | 0.574 | 2.094 | 0.445 | 2.390 | 0.328 | 2.692 | 0.230 | 2.985 | 0.147 | 3.266 | — | — | — | — |
| 14 | 1.045 | 1.350 | 0.905 | 1.551 | 0.767 | 1.779 | 0.632 | 2.030 | 0.505 | 2.296 | 0.389 | 2.572 | 0.286 | 2.848 | 0.200 | 3.111 | 0.127 | 3.360 | — | — |
| 15 | 1.077 | 1.361 | 0.946 | 1.543 | 0.814 | 1.750 | 0.685 | 1.977 | 0.562 | 2.220 | 0.447 | 2.472 | 0.343 | 2.727 | 0.251 | 2.979 | 0.175 | 3.216 | 0.111 | 3.438 |
| 16 | 1.106 | 1.371 | 0.982 | 1.539 | 0.857 | 1.728 | 0.734 | 1.935 | 0.615 | 2.157 | 0.502 | 2.388 | 0.398 | 2.624 | 0.304 | 2.860 | 0.222 | 3.090 | 0.155 | 3.304 |
| 17 | 1.133 | 1.381 | 1.015 | 1.536 | 0.897 | 1.710 | 0.779 | 1.900 | 0.664 | 2.104 | 0.554 | 2.318 | 0.451 | 2.537 | 0.356 | 2.757 | 0.272 | 2.975 | 0.198 | 3.184 |
| 18 | 1.158 | 1.391 | 1.046 | 1.535 | 0.933 | 1.696 | 0.820 | 1.872 | 0.710 | 2.060 | 0.603 | 2.257 | 0.502 | 2.461 | 0.407 | 2.667 | 0.321 | 2.873 | 0.244 | 3.073 |
| 19 | 1.180 | 1.401 | 1.074 | 1.536 | 0.967 | 1.685 | 0.859 | 1.848 | 0.752 | 2.023 | 0.649 | 2.206 | 0.549 | 2.396 | 0.456 | 2.589 | 0.369 | 2.783 | 0.290 | 2.974 |
| 20 | 1.201 | 1.411 | 1.100 | 1.537 | 0.998 | 1.676 | 0.894 | 1.828 | 0.792 | 1.991 | 0.692 | 2.162 | 0.595 | 2.339 | 0.502 | 2.521 | 0.416 | 2.704 | 0.336 | 2.885 |
| 21 | 1.221 | 1.420 | 1.125 | 1.538 | 1.026 | 1.669 | 0.927 | 1.812 | 0.829 | 1.964 | 0.732 | 2.124 | 0.637 | 2.290 | 0.547 | 2.460 | 0.461 | 2.633 | 0.380 | 2.806 |
| 22 | 1.239 | 1.429 | 1.147 | 1.541 | 1.053 | 1.664 | 0.958 | 1.797 | 0.863 | 1.940 | 0.769 | 2.090 | 0.677 | 2.246 | 0.588 | 2.407 | 0.504 | 2.571 | 0.424 | 2.734 |
| 23 | 1.257 | 1.437 | 1.168 | 1.543 | 1.078 | 1.660 | 0.986 | 1.785 | 0.895 | 1.920 | 0.804 | 2.061 | 0.715 | 2.208 | 0.628 | 2.360 | 0.545 | 2.514 | 0.465 | 2.670 |
| 24 | 1.273 | 1.446 | 1.188 | 1.546 | 1.101 | 1.656 | 1.013 | 1.775 | 0.925 | 1.902 | 0.837 | 2.035 | 0.751 | 2.174 | 0.666 | 2.318 | 0.584 | 2.464 | 0.506 | 2.613 |
| 25 | 1.288 | 1.454 | 1.206 | 1.550 | 1.123 | 1.654 | 1.038 | 1.767 | 0.953 | 1.886 | 0.868 | 2.012 | 0.784 | 2.144 | 0.702 | 2.280 | 0.621 | 2.419 | 0.544 | 2.560 |
| 26 | 1.302 | 1.461 | 1.224 | 1.553 | 1.143 | 1.652 | 1.062 | 1.759 | 0.979 | 1.873 | 0.897 | 1.992 | 0.816 | 2.117 | 0.735 | 2.246 | 0.657 | 2.379 | 0.581 | 2.513 |
| 27 | 1.316 | 1.469 | 1.240 | 1.556 | 1.162 | 1.651 | 1.084 | 1.753 | 1.004 | 1.861 | 0.925 | 1.974 | 0.845 | 2.093 | 0.767 | 2.216 | 0.691 | 2.342 | 0.616 | 2.470 |
| 28 | 1.328 | 1.476 | 1.255 | 1.560 | 1.181 | 1.650 | 1.104 | 1.747 | 1.028 | 1.850 | 0.951 | 1.958 | 0.874 | 2.071 | 0.798 | 2.188 | 0.723 | 2.309 | 0.650 | 2.431 |
| 29 | 1.341 | 1.483 | 1.270 | 1.563 | 1.198 | 1.650 | 1.124 | 1.743 | 1.050 | 1.841 | 0.975 | 1.944 | 0.900 | 2.052 | 0.826 | 2.164 | 0.753 | 2.278 | 0.682 | 2.396 |
| 30 | 1.352 | 1.489 | 1.284 | 1.567 | 1.214 | 1.650 | 1.143 | 1.739 | 1.071 | 1.833 | 0.998 | 1.931 | 0.926 | 2.034 | 0.854 | 2.141 | 0.782 | 2.251 | 0.712 | 2.363 |
| 31 | 1.363 | 1.496 | 1.297 | 1.570 | 1.229 | 1.650 | 1.160 | 1.735 | 1.090 | 1.825 | 1.020 | 1.920 | 0.950 | 2.018 | 0.879 | 2.120 | 0.810 | 2.226 | 0.741 | 2.333 |
| 32 | 1.373 | 1.502 | 1.309 | 1.574 | 1.244 | 1.650 | 1.177 | 1.732 | 1.109 | 1.819 | 1.041 | 1.909 | 0.972 | 2.004 | 0.904 | 2.102 | 0.836 | 2.203 | 0.769 | 2.306 |
| 33 | 1.383 | 1.508 | 1.321 | 1.577 | 1.258 | 1.651 | 1.193 | 1.730 | 1.127 | 1.813 | 1.061 | 1.900 | 0.994 | 1.991 | 0.927 | 2.085 | 0.861 | 2.181 | 0.795 | 2.281 |
| 34 | 1.393 | 1.514 | 1.333 | 1.580 | 1.271 | 1.652 | 1.208 | 1.728 | 1.144 | 1.808 | 1.080 | 1.891 | 1.015 | 1.979 | 0.950 | 2.069 | 0.885 | 2.162 | 0.821 | 2.257 |
| 35 | 1.402 | 1.519 | 1.343 | 1.584 | 1.283 | 1.653 | 1.222 | 1.726 | 1.160 | 1.803 | 1.097 | 1.884 | 1.034 | 1.967 | 0.971 | 2.054 | 0.908 | 2.144 | 0.845 | 2.236 |
| 36 | 1.411 | 1.525 | 1.354 | 1.587 | 1.295 | 1.654 | 1.236 | 1.724 | 1.175 | 1.799 | 1.114 | 1.877 | 1.053 | 1.957 | 0.991 | 2.041 | 0.930 | 2.127 | 0.868 | 2.216 |
| 37 | 1.419 | 1.530 | 1.364 | 1.590 | 1.307 | 1.655 | 1.249 | 1.723 | 1.190 | 1.795 | 1.131 | 1.870 | 1.071 | 1.948 | 1.011 | 2.029 | 0.951 | 2.112 | 0.891 | 2.198 |
| 38 | 1.427 | 1.535 | 1.373 | 1.594 | 1.318 | 1.656 | 1.261 | 1.722 | 1.204 | 1.792 | 1.146 | 1.864 | 1.088 | 1.939 | 1.029 | 2.017 | 0.970 | 2.098 | 0.912 | 2.180 |
| 39 | 1.435 | 1.540 | 1.382 | 1.597 | 1.328 | 1.658 | 1.273 | 1.722 | 1.218 | 1.789 | 1.161 | 1.859 | 1.104 | 1.932 | 1.047 | 2.007 | 0.990 | 2.085 | 0.932 | 2.164 |
| 40 | 1.442 | 1.544 | 1.391 | 1.600 | 1.338 | 1.659 | 1.285 | 1.721 | 1.230 | 1.786 | 1.175 | 1.854 | 1.120 | 1.924 | 1.064 | 1.997 | 1.008 | 2.072 | 0.952 | 2.149 |
| 45 | 1.475 | 1.566 | 1.430 | 1.615 | 1.383 | 1.666 | 1.336 | 1.720 | 1.287 | 1.776 | 1.238 | 1.835 | 1.189 | 1.895 | 1.139 | 1.958 | 1.089 | 2.022 | 1.038 | 2.088 |
| 50 | 1.503 | 1.585 | 1.462 | 1.628 | 1.421 | 1.674 | 1.378 | 1.721 | 1.335 | 1.771 | 1.291 | 1.822 | 1.246 | 1.875 | 1.201 | 1.930 | 1.156 | 1.986 | 1.110 | 2.044 |
| 55 | 1.528 | 1.601 | 1.490 | 1.641 | 1.452 | 1.681 | 1.414 | 1.724 | 1.374 | 1.768 | 1.334 | 1.814 | 1.294 | 1.861 | 1.253 | 1.909 | 1.212 | 1.959 | 1.170 | 2.010 |
| 60 | 1.549 | 1.616 | 1.514 | 1.652 | 1.480 | 1.689 | 1.444 | 1.727 | 1.408 | 1.767 | 1.372 | 1.808 | 1.335 | 1.850 | 1.298 | 1.894 | 1.260 | 1.939 | 1.222 | 1.984 |
| 65 | 1.567 | 1.629 | 1.536 | 1.662 | 1.503 | 1.696 | 1.471 | 1.731 | 1.438 | 1.767 | 1.404 | 1.805 | 1.370 | 1.843 | 1.336 | 1.882 | 1.301 | 1.923 | 1.266 | 1.964 |
| 70 | 1.583 | 1.641 | 1.554 | 1.672 | 1.525 | 1.703 | 1.494 | 1.735 | 1.464 | 1.768 | 1.433 | 1.802 | 1.401 | 1.837 | 1.369 | 1.873 | 1.337 | 1.910 | 1.305 | 1.948 |
| 75 | 1.598 | 1.652 | 1.571 | 1.680 | 1.543 | 1.709 | 1.515 | 1.739 | 1.487 | 1.770 | 1.458 | 1.801 | 1.428 | 1.834 | 1.399 | 1.867 | 1.369 | 1.901 | 1.339 | 1.935 |
| 80 | 1.611 | 1.662 | 1.586 | 1.688 | 1.560 | 1.715 | 1.534 | 1.743 | 1.507 | 1.772 | 1.480 | 1.801 | 1.453 | 1.831 | 1.425 | 1.861 | 1.397 | 1.893 | 1.369 | 1.925 |
| 85 | 1.624 | 1.671 | 1.600 | 1.696 | 1.575 | 1.721 | 1.550 | 1.747 | 1.525 | 1.774 | 1.500 | 1.801 | 1.474 | 1.829 | 1.448 | 1.857 | 1.422 | 1.886 | 1.396 | 1.916 |
| 90 | 1.635 | 1.679 | 1.612 | 1.703 | 1.589 | 1.726 | 1.566 | 1.751 | 1.542 | 1.776 | 1.518 | 1.801 | 1.494 | 1.827 | 1.469 | 1.854 | 1.445 | 1.881 | 1.420 | 1.909 |
| 95 | 1.645 | 1.687 | 1.623 | 1.709 | 1.602 | 1.732 | 1.579 | 1.755 | 1.557 | 1.778 | 1.535 | 1.802 | 1.512 | 1.827 | 1.489 | 1.852 | 1.465 | 1.877 | 1.442 | 1.903 |
| 100 | 1.654 | 1.694 | 1.634 | 1.715 | 1.613 | 1.736 | 1.592 | 1.758 | 1.571 | 1.780 | 1.550 | 1.803 | 1.528 | 1.826 | 1.506 | 1.850 | 1.484 | 1.874 | 1.462 | 1.898 |
| 150 | 1.720 | 1.746 | 1.706 | 1.760 | 1.693 | 1.774 | 1.679 | 1.788 | 1.665 | 1.802 | 1.651 | 1.817 | 1.637 | 1.832 | 1.622 | 1.847 | 1.608 | 1.862 | 1.594 | 1.877 |
| 200 | 1.758 | 1.778 | 1.748 | 1.789 | 1.738 | 1.799 | 1.728 | 1.810 | 1.718 | 1.820 | 1.707 | 1.831 | 1.697 | 1.841 | 1.686 | 1.852 | 1.675 | 1.863 | 1.665 | 1.874 |

(*continúa*)

(*continuación*)

| | $k'=11$ | | $k'=12$ | | $k'=13$ | | $k'=14$ | | $k'=15$ | | $k'=16$ | | $k'=17$ | | $k'=18$ | | $k'=19$ | | $k'=20$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 16 | 0.098 | 3.503 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 17 | 0.138 | 3.378 | 0.087 | 3.557 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 18 | 0.177 | 3.265 | 0.123 | 3.441 | 0.078 | 3.603 | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 19 | 0.220 | 3.159 | 0.160 | 3.335 | 0.111 | 3.496 | 0.070 | 3.642 | — | — | — | — | — | — | — | — | — | — | — | — |
| 20 | 0.263 | 3.063 | 0.200 | 3.234 | 0.145 | 3.395 | 0.100 | 3.542 | 0.063 | 3.676 | — | — | — | — | — | — | — | — | — | — |
| 21 | 0.307 | 2.976 | 0.240 | 3.141 | 0.182 | 3.300 | 0.132 | 3.448 | 0.091 | 3.583 | 0.058 | 3.705 | — | — | — | — | — | — | — | — |
| 22 | 0.349 | 2.897 | 0.281 | 3.057 | 0.220 | 3.211 | 0.166 | 3.358 | 0.120 | 3.495 | 0.083 | 3.619 | 0.052 | 3.731 | — | — | — | — | — | — |
| 23 | 0.391 | 2.826 | 0.322 | 2.979 | 0.259 | 3.128 | 0.202 | 3.272 | 0.153 | 3.409 | 0.110 | 3.535 | 0.076 | 3.650 | 0.048 | 3.753 | — | — | — | — |
| 24 | 0.431 | 2.761 | 0.362 | 2.908 | 0.297 | 3.053 | 0.239 | 3.193 | 0.186 | 3.327 | 0.141 | 3.454 | 0.101 | 3.572 | 0.070 | 3.678 | 0.044 | 3.773 | — | — |
| 25 | 0.470 | 2.702 | 0.400 | 2.844 | 0.335 | 2.983 | 0.275 | 3.119 | 0.221 | 3.251 | 0.172 | 3.376 | 0.130 | 3.494 | 0.094 | 3.604 | 0.065 | 3.702 | 0.041 | 3.790 |
| 26 | 0.508 | 2.649 | 0.438 | 2.784 | 0.373 | 2.919 | 0.312 | 3.051 | 0.256 | 3.179 | 0.205 | 3.303 | 0.160 | 3.420 | 0.120 | 3.531 | 0.087 | 3.632 | 0.060 | 3.724 |
| 27 | 0.544 | 2.600 | 0.475 | 2.730 | 0.409 | 2.859 | 0.348 | 2.987 | 0.291 | 3.112 | 0.238 | 3.233 | 0.191 | 3.349 | 0.149 | 3.460 | 0.112 | 3.563 | 0.081 | 3.658 |
| 28 | 0.578 | 2.555 | 0.510 | 2.680 | 0.445 | 2.805 | 0.383 | 2.928 | 0.325 | 3.050 | 0.271 | 3.168 | 0.222 | 3.283 | 0.178 | 3.392 | 0.138 | 3.495 | 0.104 | 3.592 |
| 29 | 0.612 | 2.515 | 0.544 | 2.634 | 0.479 | 2.755 | 0.418 | 2.874 | 0.359 | 2.992 | 0.305 | 3.107 | 0.254 | 3.219 | 0.208 | 3.327 | 0.166 | 3.431 | 0.129 | 3.528 |
| 30 | 0.643 | 2.477 | 0.577 | 2.592 | 0.512 | 2.708 | 0.451 | 2.823 | 0.392 | 2.937 | 0.337 | 3.050 | 0.286 | 3.160 | 0.238 | 3.266 | 0.195 | 3.368 | 0.156 | 3.465 |
| 31 | 0.674 | 2.443 | 0.608 | 2.553 | 0.545 | 2.665 | 0.484 | 2.776 | 0.425 | 2.887 | 0.370 | 2.996 | 0.317 | 3.103 | 0.269 | 3.208 | 0.224 | 3.309 | 0.183 | 3.406 |
| 32 | 0.703 | 2.411 | 0.638 | 2.517 | 0.576 | 2.625 | 0.515 | 2.733 | 0.457 | 2.840 | 0.401 | 2.946 | 0.349 | 3.050 | 0.299 | 3.153 | 0.253 | 3.252 | 0.211 | 3.348 |
| 33 | 0.731 | 2.382 | 0.668 | 2.484 | 0.606 | 2.588 | 0.546 | 2.692 | 0.488 | 2.796 | 0.432 | 2.899 | 0.379 | 3.000 | 0.329 | 3.100 | 0.283 | 3.198 | 0.239 | 3.293 |
| 34 | 0.758 | 2.355 | 0.695 | 2.454 | 0.634 | 2.554 | 0.575 | 2.654 | 0.518 | 2.754 | 0.462 | 2.854 | 0.409 | 2.954 | 0.359 | 3.051 | 0.312 | 3.147 | 0.267 | 3.240 |
| 35 | 0.783 | 2.330 | 0.722 | 2.425 | 0.662 | 2.521 | 0.604 | 2.619 | 0.547 | 2.716 | 0.492 | 2.813 | 0.439 | 2.910 | 0.388 | 3.005 | 0.340 | 3.099 | 0.295 | 3.190 |
| 36 | 0.808 | 2.306 | 0.748 | 2.398 | 0.689 | 2.492 | 0.631 | 2.586 | 0.575 | 2.680 | 0.520 | 2.774 | 0.467 | 2.868 | 0.417 | 2.961 | 0.369 | 3.053 | 0.323 | 3.142 |
| 37 | 0.831 | 2.285 | 0.772 | 2.374 | 0.714 | 2.464 | 0.657 | 2.555 | 0.602 | 2.646 | 0.548 | 2.738 | 0.495 | 2.829 | 0.445 | 2.920 | 0.397 | 3.009 | 0.351 | 3.097 |
| 38 | 0.854 | 2.265 | 0.796 | 2.351 | 0.739 | 2.438 | 0.683 | 2.526 | 0.628 | 2.614 | 0.575 | 2.703 | 0.522 | 2.792 | 0.472 | 2.880 | 0.424 | 2.968 | 0.378 | 3.054 |
| 39 | 0.875 | 2.246 | 0.819 | 2.329 | 0.763 | 2.413 | 0.707 | 2.499 | 0.653 | 2.585 | 0.600 | 2.671 | 0.549 | 2.757 | 0.499 | 2.843 | 0.451 | 2.929 | 0.404 | 3.013 |
| 40 | 0.896 | 2.228 | 0.840 | 2.309 | 0.785 | 2.391 | 0.731 | 2.473 | 0.678 | 2.557 | 0.626 | 2.641 | 0.575 | 2.724 | 0.525 | 2.808 | 0.477 | 2.892 | 0.430 | 2.974 |
| 45 | 0.988 | 2.156 | 0.938 | 2.225 | 0.887 | 2.296 | 0.838 | 2.367 | 0.788 | 2.439 | 0.740 | 2.512 | 0.692 | 2.586 | 0.644 | 2.659 | 0.598 | 2.733 | 0.553 | 2.807 |
| 50 | 1.064 | 2.103 | 1.019 | 2.163 | 0.973 | 2.225 | 0.927 | 2.287 | 0.882 | 2.350 | 0.836 | 2.414 | 0.792 | 2.479 | 0.747 | 2.544 | 0.703 | 2.610 | 0.660 | 2.675 |
| 55 | 1.129 | 2.062 | 1.087 | 2.116 | 1.045 | 2.170 | 1.003 | 2.225 | 0.961 | 2.281 | 0.919 | 2.338 | 0.877 | 2.396 | 0.836 | 2.454 | 0.795 | 2.512 | 0.754 | 2.571 |
| 60 | 1.184 | 2.031 | 1.145 | 2.079 | 1.106 | 2.127 | 1.068 | 2.177 | 1.029 | 2.227 | 0.990 | 2.278 | 0.951 | 2.330 | 0.913 | 2.382 | 0.874 | 2.434 | 0.836 | 2.487 |
| 65 | 1.231 | 2.006 | 1.195 | 2.049 | 1.160 | 2.093 | 1.124 | 2.138 | 1.088 | 2.183 | 1.052 | 2.229 | 1.016 | 2.276 | 0.980 | 2.323 | 0.944 | 2.371 | 0.908 | 2.419 |
| 70 | 1.272 | 1.986 | 1.239 | 2.026 | 1.206 | 2.066 | 1.172 | 2.106 | 1.139 | 2.148 | 1.105 | 2.189 | 1.072 | 2.232 | 1.038 | 2.275 | 1.005 | 2.318 | 0.971 | 2.362 |
| 75 | 1.308 | 1.970 | 1.277 | 2.006 | 1.247 | 2.043 | 1.215 | 2.080 | 1.184 | 2.118 | 1.153 | 2.156 | 1.121 | 2.195 | 1.090 | 2.235 | 1.058 | 2.275 | 1.027 | 2.315 |
| 80 | 1.340 | 1.957 | 1.311 | 1.991 | 1.283 | 2.024 | 1.253 | 2.059 | 1.224 | 2.093 | 1.195 | 2.129 | 1.165 | 2.165 | 1.136 | 2.201 | 1.106 | 2.238 | 1.076 | 2.275 |
| 85 | 1.369 | 1.946 | 1.342 | 1.977 | 1.315 | 2.009 | 1.287 | 2.040 | 1.260 | 2.073 | 1.232 | 2.105 | 1.205 | 2.139 | 1.177 | 2.172 | 1.149 | 2.206 | 1.121 | 2.241 |
| 90 | 1.395 | 1.937 | 1.369 | 1.966 | 1.344 | 1.995 | 1.318 | 2.025 | 1.292 | 2.055 | 1.266 | 2.085 | 1.240 | 2.116 | 1.213 | 2.148 | 1.187 | 2.179 | 1.160 | 2.211 |
| 95 | 1.418 | 1.929 | 1.394 | 1.956 | 1.370 | 1.984 | 1.345 | 2.012 | 1.321 | 2.040 | 1.296 | 2.068 | 1.271 | 2.097 | 1.247 | 2.126 | 1.222 | 2.156 | 1.197 | 2.186 |
| 100 | 1.439 | 1.923 | 1.416 | 1.948 | 1.393 | 1.974 | 1.371 | 2.000 | 1.347 | 2.026 | 1.324 | 2.053 | 1.301 | 2.080 | 1.277 | 2.108 | 1.253 | 2.135 | 1.229 | 2.164 |
| 150 | 1.579 | 1.892 | 1.564 | 1.908 | 1.550 | 1.924 | 1.535 | 1.940 | 1.519 | 1.956 | 1.504 | 1.972 | 1.489 | 1.989 | 1.474 | 2.006 | 1.458 | 2.023 | 1.443 | 2.040 |
| 200 | 1.654 | 1.885 | 1.643 | 1.896 | 1.632 | 1.908 | 1.621 | 1.919 | 1.610 | 1.931 | 1.599 | 1.943 | 1.588 | 1.955 | 1.576 | 1.967 | 1.565 | 1.979 | 1.554 | 1.991 |

*Nota:* $n$ = número de observaciones, $k'$ = número de variables explicativas, excluyendo el término constante.

Fuente: Esta tabla es una extensión de la tabla original de Durbin-Watson y se reproduce de N.E. Savin y K.J. White, "The Durbin-Watson Test for Serial Correlation with Extreme Small Samples or Many Regressors", *Econometrica*, vol. 45, noviembre de 1977, pp. 1989-1996, con las correcciones de R.W. Farebrother, *Econometrica*, vol. 48, septiembre de 1980, p. 1554. Se reproduce con autorización de la Sociedad Econométrica.

**EJEMPLO 1**

Si $n = 40$ y $k' = 4$, $d_L = 1.285$ y $d_U = 1.721$. Si un valor $d$ calculado es menor que 1.285, hay evidencia de correlación serial positiva de primer orden; si es mayor que 1.721, no hay evidencia de correlación serial positiva de primer orden; pero si $d$ se encuentra entre el límite inferior y el superior, hay evidencia inconclusa relacionada con la presencia o ausencia de correlación serial positiva de primer orden.

**TABLA D.5B** **Estadístico *d* de Durbin-Watson: puntos de significancia de $d_L$ y $d_U$ en el nivel de significancia de 0.01**

| | $k' = 1$ | | $k' = 2$ | | $k' = 3$ | | $k' = 4$ | | $k' = 5$ | | $k' = 6$ | | $k' = 7$ | | $k' = 8$ | | $k' = 9$ | | $k' = 10$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 6 | 0.390 | 1.142 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 7 | 0.435 | 1.036 | 0.294 | 1.676 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 8 | 0.497 | 1.003 | 0.345 | 1.489 | 0.229 | 2.102 | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 9 | 0.554 | 0.998 | 0.408 | 1.389 | 0.279 | 1.875 | 0.183 | 2.433 | — | — | — | — | — | — | — | — | — | — | — | — |
| 10 | 0.604 | 1.001 | 0.466 | 1.333 | 0.340 | 1.733 | 0.230 | 2.193 | 0.150 | 2.690 | — | — | — | — | — | — | — | — | — | — |
| 11 | 0.653 | 1.010 | 0.519 | 1.297 | 0.396 | 1.640 | 0.286 | 2.030 | 0.193 | 2.453 | 0.124 | 2.892 | — | — | — | — | — | — | — | — |
| 12 | 0.697 | 1.023 | 0.569 | 1.274 | 0.449 | 1.575 | 0.339 | 1.913 | 0.244 | 2.280 | 0.164 | 2.665 | 0.105 | 3.053 | — | — | — | — | — | — |
| 13 | 0.738 | 1.038 | 0.616 | 1.261 | 0.499 | 1.526 | 0.391 | 1.826 | 0.294 | 2.150 | 0.211 | 2.490 | 0.140 | 2.838 | 0.090 | 3.182 | — | — | — | — |
| 14 | 0.776 | 1.054 | 0.660 | 1.254 | 0.547 | 1.490 | 0.441 | 1.757 | 0.343 | 2.049 | 0.257 | 2.354 | 0.183 | 2.667 | 0.122 | 2.981 | 0.078 | 3.287 | — | — |
| 15 | 0.811 | 1.070 | 0.700 | 1.252 | 0.591 | 1.464 | 0.488 | 1.704 | 0.391 | 1.967 | 0.303 | 2.244 | 0.226 | 2.530 | 0.161 | 2.817 | 0.107 | 3.101 | 0.068 | 3.374 |
| 16 | 0.844 | 1.086 | 0.737 | 1.252 | 0.633 | 1.446 | 0.532 | 1.663 | 0.437 | 1.900 | 0.349 | 2.153 | 0.269 | 2.416 | 0.200 | 2.681 | 0.142 | 2.944 | 0.094 | 3.201 |
| 17 | 0.874 | 1.102 | 0.772 | 1.255 | 0.672 | 1.432 | 0.574 | 1.630 | 0.480 | 1.847 | 0.393 | 2.078 | 0.313 | 2.319 | 0.241 | 2.566 | 0.179 | 2.811 | 0.127 | 3.053 |
| 18 | 0.902 | 1.118 | 0.805 | 1.259 | 0.708 | 1.422 | 0.613 | 1.604 | 0.522 | 1.803 | 0.435 | 2.015 | 0.355 | 2.238 | 0.282 | 2.467 | 0.216 | 2.697 | 0.160 | 2.925 |
| 19 | 0.928 | 1.132 | 0.835 | 1.265 | 0.742 | 1.415 | 0.650 | 1.584 | 0.561 | 1.767 | 0.476 | 1.963 | 0.396 | 2.169 | 0.322 | 2.381 | 0.255 | 2.597 | 0.196 | 2.813 |
| 20 | 0.952 | 1.147 | 0.863 | 1.271 | 0.773 | 1.411 | 0.685 | 1.567 | 0.598 | 1.737 | 0.515 | 1.918 | 0.436 | 2.110 | 0.362 | 2.308 | 0.294 | 2.510 | 0.232 | 2.714 |
| 21 | 0.975 | 1.161 | 0.890 | 1.277 | 0.803 | 1.408 | 0.718 | 1.554 | 0.633 | 1.712 | 0.552 | 1.881 | 0.474 | 2.059 | 0.400 | 2.244 | 0.331 | 2.434 | 0.268 | 2.625 |
| 22 | 0.997 | 1.174 | 0.914 | 1.284 | 0.831 | 1.407 | 0.748 | 1.543 | 0.667 | 1.691 | 0.587 | 1.849 | 0.510 | 2.015 | 0.437 | 2.188 | 0.368 | 2.367 | 0.304 | 2.548 |
| 23 | 1.018 | 1.187 | 0.938 | 1.291 | 0.858 | 1.407 | 0.777 | 1.534 | 0.698 | 1.673 | 0.620 | 1.821 | 0.545 | 1.977 | 0.473 | 2.140 | 0.404 | 2.308 | 0.340 | 2.479 |
| 24 | 1.037 | 1.199 | 0.960 | 1.298 | 0.882 | 1.407 | 0.805 | 1.528 | 0.728 | 1.658 | 0.652 | 1.797 | 0.578 | 1.944 | 0.507 | 2.097 | 0.439 | 2.255 | 0.375 | 2.417 |
| 25 | 1.055 | 1.211 | 0.981 | 1.305 | 0.906 | 1.409 | 0.831 | 1.523 | 0.756 | 1.645 | 0.682 | 1.776 | 0.610 | 1.915 | 0.540 | 2.059 | 0.473 | 2.209 | 0.409 | 2.362 |
| 26 | 1.072 | 1.222 | 1.001 | 1.312 | 0.928 | 1.411 | 0.855 | 1.518 | 0.783 | 1.635 | 0.711 | 1.759 | 0.640 | 1.889 | 0.572 | 2.026 | 0.505 | 2.168 | 0.441 | 2.313 |
| 27 | 1.089 | 1.233 | 1.019 | 1.319 | 0.949 | 1.413 | 0.878 | 1.515 | 0.808 | 1.626 | 0.738 | 1.743 | 0.669 | 1.867 | 0.602 | 1.997 | 0.536 | 2.131 | 0.473 | 2.269 |
| 28 | 1.104 | 1.244 | 1.037 | 1.325 | 0.969 | 1.415 | 0.900 | 1.513 | 0.832 | 1.618 | 0.764 | 1.729 | 0.696 | 1.847 | 0.630 | 1.970 | 0.566 | 2.098 | 0.504 | 2.229 |
| 29 | 1.119 | 1.254 | 1.054 | 1.332 | 0.988 | 1.418 | 0.921 | 1.512 | 0.855 | 1.611 | 0.788 | 1.718 | 0.723 | 1.830 | 0.658 | 1.947 | 0.595 | 2.068 | 0.533 | 2.193 |
| 30 | 1.133 | 1.263 | 1.070 | 1.339 | 1.006 | 1.421 | 0.941 | 1.511 | 0.877 | 1.606 | 0.812 | 1.707 | 0.748 | 1.814 | 0.684 | 1.925 | 0.622 | 2.041 | 0.562 | 2.160 |
| 31 | 1.147 | 1.273 | 1.085 | 1.345 | 1.023 | 1.425 | 0.960 | 1.510 | 0.897 | 1.601 | 0.834 | 1.698 | 0.772 | 1.800 | 0.710 | 1.906 | 0.649 | 2.017 | 0.589 | 2.131 |
| 32 | 1.160 | 1.282 | 1.100 | 1.352 | 1.040 | 1.428 | 0.979 | 1.510 | 0.917 | 1.597 | 0.856 | 1.690 | 0.794 | 1.788 | 0.734 | 1.889 | 0.674 | 1.995 | 0.615 | 2.104 |
| 33 | 1.172 | 1.291 | 1.114 | 1.358 | 1.055 | 1.432 | 0.996 | 1.510 | 0.936 | 1.594 | 0.876 | 1.683 | 0.816 | 1.776 | 0.757 | 1.874 | 0.698 | 1.975 | 0.641 | 2.080 |
| 34 | 1.184 | 1.299 | 1.128 | 1.364 | 1.070 | 1.435 | 1.012 | 1.511 | 0.954 | 1.591 | 0.896 | 1.677 | 0.837 | 1.766 | 0.779 | 1.860 | 0.722 | 1.957 | 0.665 | 2.057 |
| 35 | 1.195 | 1.307 | 1.140 | 1.370 | 1.085 | 1.439 | 1.028 | 1.512 | 0.971 | 1.589 | 0.914 | 1.671 | 0.857 | 1.757 | 0.800 | 1.847 | 0.744 | 1.940 | 0.689 | 2.037 |
| 36 | 1.206 | 1.315 | 1.153 | 1.376 | 1.098 | 1.442 | 1.043 | 1.513 | 0.988 | 1.588 | 0.932 | 1.666 | 0.877 | 1.749 | 0.821 | 1.836 | 0.766 | 1.925 | 0.711 | 2.018 |
| 37 | 1.217 | 1.323 | 1.165 | 1.382 | 1.112 | 1.446 | 1.058 | 1.514 | 1.004 | 1.586 | 0.950 | 1.662 | 0.895 | 1.742 | 0.841 | 1.825 | 0.787 | 1.911 | 0.733 | 2.001 |
| 38 | 1.227 | 1.330 | 1.176 | 1.388 | 1.124 | 1.449 | 1.072 | 1.515 | 1.019 | 1.585 | 0.966 | 1.658 | 0.913 | 1.735 | 0.860 | 1.816 | 0.807 | 1.899 | 0.754 | 1.985 |
| 39 | 1.237 | 1.337 | 1.187 | 1.393 | 1.137 | 1.453 | 1.085 | 1.517 | 1.034 | 1.584 | 0.982 | 1.655 | 0.930 | 1.729 | 0.878 | 1.807 | 0.826 | 1.887 | 0.774 | 1.970 |
| 40 | 1.246 | 1.344 | 1.198 | 1.398 | 1.148 | 1.457 | 1.098 | 1.518 | 1.048 | 1.584 | 0.997 | 1.652 | 0.946 | 1.724 | 0.895 | 1.799 | 0.844 | 1.876 | 0.749 | 1.956 |
| 45 | 1.288 | 1.376 | 1.245 | 1.423 | 1.201 | 1.474 | 1.156 | 1.528 | 1.111 | 1.584 | 1.065 | 1.643 | 1.019 | 1.704 | 0.974 | 1.768 | 0.927 | 1.834 | 0.881 | 1.902 |
| 50 | 1.324 | 1.403 | 1.285 | 1.446 | 1.245 | 1.491 | 1.205 | 1.538 | 1.164 | 1.587 | 1.123 | 1.639 | 1.081 | 1.692 | 1.039 | 1.748 | 0.997 | 1.805 | 0.955 | 1.864 |
| 55 | 1.356 | 1.427 | 1.320 | 1.466 | 1.284 | 1.506 | 1.247 | 1.548 | 1.209 | 1.592 | 1.172 | 1.638 | 1.134 | 1.685 | 1.095 | 1.734 | 1.057 | 1.785 | 1.018 | 1.837 |
| 60 | 1.383 | 1.449 | 1.350 | 1.484 | 1.317 | 1.520 | 1.283 | 1.558 | 1.249 | 1.598 | 1.214 | 1.639 | 1.179 | 1.682 | 1.144 | 1.726 | 1.108 | 1.771 | 1.072 | 1.817 |
| 65 | 1.407 | 1.468 | 1.377 | 1.500 | 1.346 | 1.534 | 1.315 | 1.568 | 1.283 | 1.604 | 1.251 | 1.642 | 1.218 | 1.680 | 1.186 | 1.720 | 1.153 | 1.761 | 1.120 | 1.802 |
| 70 | 1.429 | 1.485 | 1.400 | 1.515 | 1.372 | 1.546 | 1.343 | 1.578 | 1.313 | 1.611 | 1.283 | 1.645 | 1.253 | 1.680 | 1.223 | 1.716 | 1.192 | 1.754 | 1.162 | 1.792 |
| 75 | 1.448 | 1.501 | 1.422 | 1.529 | 1.395 | 1.557 | 1.368 | 1.587 | 1.340 | 1.617 | 1.313 | 1.649 | 1.284 | 1.682 | 1.256 | 1.714 | 1.227 | 1.748 | 1.199 | 1.783 |
| 80 | 1.466 | 1.515 | 1.441 | 1.541 | 1.416 | 1.568 | 1.390 | 1.595 | 1.364 | 1.624 | 1.338 | 1.653 | 1.312 | 1.683 | 1.285 | 1.714 | 1.259 | 1.745 | 1.232 | 1.777 |
| 85 | 1.482 | 1.528 | 1.458 | 1.553 | 1.435 | 1.578 | 1.411 | 1.603 | 1.386 | 1.630 | 1.362 | 1.657 | 1.337 | 1.685 | 1.312 | 1.714 | 1.287 | 1.743 | 1.262 | 1.773 |
| 90 | 1.496 | 1.540 | 1.474 | 1.563 | 1.452 | 1.587 | 1.429 | 1.611 | 1.406 | 1.636 | 1.383 | 1.661 | 1.360 | 1.687 | 1.336 | 1.714 | 1.312 | 1.741 | 1.288 | 1.769 |
| 95 | 1.510 | 1.552 | 1.489 | 1.573 | 1.468 | 1.596 | 1.446 | 1.618 | 1.425 | 1.642 | 1.403 | 1.666 | 1.381 | 1.690 | 1.358 | 1.715 | 1.336 | 1.741 | 1.313 | 1.767 |
| 100 | 1.522 | 1.562 | 1.503 | 1.583 | 1.482 | 1.604 | 1.462 | 1.625 | 1.441 | 1.647 | 1.421 | 1.670 | 1.400 | 1.693 | 1.378 | 1.717 | 1.357 | 1.741 | 1.335 | 1.765 |
| 150 | 1.611 | 1.637 | 1.598 | 1.651 | 1.584 | 1.665 | 1.571 | 1.679 | 1.557 | 1.693 | 1.543 | 1.708 | 1.530 | 1.722 | 1.515 | 1.737 | 1.501 | 1.752 | 1.486 | 1.767 |
| 200 | 1.664 | 1.684 | 1.653 | 1.693 | 1.643 | 1.704 | 1.633 | 1.715 | 1.623 | 1.725 | 1.613 | 1.735 | 1.603 | 1.746 | 1.592 | 1.757 | 1.582 | 1.768 | 1.571 | 1.779 |

(*continúa*)

(*continuación*)

| | $k'=11$ | | $k'=12$ | | $k'=13$ | | $k'=14$ | | $k'=15$ | | $k'=16$ | | $k'=17$ | | $k'=18$ | | $k'=19$ | | $k'=20$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 16 | 0.060 | 3.446 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 17 | 0.084 | 3.286 | 0.053 | 3.506 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 18 | 0.113 | 3.146 | 0.075 | 3.358 | 0.047 | 3.357 | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| 19 | 0.145 | 3.023 | 0.102 | 3.227 | 0.067 | 3.420 | 0.043 | 3.601 | — | — | — | — | — | — | — | — | — | — | — | — |
| 20 | 0.178 | 2.914 | 0.131 | 3.109 | 0.092 | 3.297 | 0.061 | 3.474 | 0.038 | 3.639 | — | — | — | — | — | — | — | — | — | — |
| 21 | 0.212 | 2.817 | 0.162 | 3.004 | 0.119 | 3.185 | 0.084 | 3.358 | 0.055 | 3.521 | 0.035 | 3.671 | — | — | — | — | — | — | — | — |
| 22 | 0.246 | 2.729 | 0.194 | 2.909 | 0.148 | 3.084 | 0.109 | 3.252 | 0.077 | 3.412 | 0.050 | 3.562 | 0.032 | 3.700 | — | — | — | — | — | — |
| 23 | 0.281 | 2.651 | 0.227 | 2.822 | 0.178 | 2.991 | 0.136 | 3.155 | 0.100 | 3.311 | 0.070 | 3.459 | 0.046 | 3.597 | 0.029 | 3.725 | — | — | — | — |
| 24 | 0.315 | 2.580 | 0.260 | 2.744 | 0.209 | 2.906 | 0.165 | 3.065 | 0.125 | 3.218 | 0.092 | 3.363 | 0.065 | 3.501 | 0.043 | 3.629 | 0.027 | 3.747 | — | — |
| 25 | 0.348 | 2.517 | 0.292 | 2.674 | 0.240 | 2.829 | 0.194 | 2.982 | 0.152 | 3.131 | 0.116 | 3.274 | 0.085 | 3.410 | 0.060 | 3.538 | 0.039 | 3.657 | 0.025 | 3.766 |
| 26 | 0.381 | 2.460 | 0.324 | 2.610 | 0.272 | 2.758 | 0.224 | 2.906 | 0.180 | 3.050 | 0.141 | 3.191 | 0.107 | 3.325 | 0.079 | 3.452 | 0.055 | 3.572 | 0.036 | 3.682 |
| 27 | 0.413 | 2.409 | 0.356 | 2.552 | 0.303 | 2.694 | 0.253 | 2.836 | 0.208 | 2.976 | 0.167 | 3.113 | 0.131 | 3.245 | 0.100 | 3.371 | 0.073 | 3.490 | 0.051 | 3.602 |
| 28 | 0.444 | 2.363 | 0.387 | 2.499 | 0.333 | 2.635 | 0.283 | 2.772 | 0.237 | 2.907 | 0.194 | 3.040 | 0.156 | 3.169 | 0.122 | 3.294 | 0.093 | 3.412 | 0.068 | 3.524 |
| 29 | 0.474 | 2.321 | 0.417 | 2.451 | 0.363 | 2.582 | 0.313 | 2.713 | 0.266 | 2.843 | 0.222 | 2.972 | 0.182 | 3.098 | 0.146 | 3.220 | 0.114 | 3.338 | 0.087 | 3.450 |
| 30 | 0.503 | 2.283 | 0.447 | 2.407 | 0.393 | 2.533 | 0.342 | 2.659 | 0.294 | 2.785 | 0.249 | 2.909 | 0.208 | 3.032 | 0.171 | 3.152 | 0.137 | 3.267 | 0.107 | 3.379 |
| 31 | 0.531 | 2.248 | 0.475 | 2.367 | 0.422 | 2.487 | 0.371 | 2.609 | 0.322 | 2.730 | 0.277 | 2.851 | 0.234 | 2.970 | 0.196 | 3.087 | 0.160 | 3.201 | 0.128 | 3.311 |
| 32 | 0.558 | 2.216 | 0.503 | 2.330 | 0.450 | 2.446 | 0.399 | 2.563 | 0.350 | 2.680 | 0.304 | 2.797 | 0.261 | 2.912 | 0.221 | 3.026 | 0.184 | 3.137 | 0.151 | 3.246 |
| 33 | 0.585 | 2.187 | 0.530 | 2.296 | 0.477 | 2.408 | 0.426 | 2.520 | 0.377 | 2.633 | 0.331 | 2.746 | 0.287 | 2.858 | 0.246 | 2.969 | 0.209 | 3.078 | 0.174 | 3.184 |
| 34 | 0.610 | 2.160 | 0.556 | 2.266 | 0.503 | 2.373 | 0.452 | 2.481 | 0.404 | 2.590 | 0.357 | 2.699 | 0.313 | 2.808 | 0.272 | 2.915 | 0.233 | 3.022 | 0.197 | 3.126 |
| 35 | 0.634 | 2.136 | 0.581 | 2.237 | 0.529 | 2.340 | 0.478 | 2.444 | 0.430 | 2.550 | 0.383 | 2.655 | 0.339 | 2.761 | 0.297 | 2.865 | 0.257 | 2.969 | 0.221 | 3.071 |
| 36 | 0.658 | 2.113 | 0.605 | 2.210 | 0.554 | 2.310 | 0.504 | 2.410 | 0.455 | 2.512 | 0.409 | 2.614 | 0.364 | 2.717 | 0.322 | 2.818 | 0.282 | 2.919 | 0.244 | 3.019 |
| 37 | 0.680 | 2.092 | 0.628 | 2.186 | 0.578 | 2.282 | 0.528 | 2.379 | 0.480 | 2.477 | 0.434 | 2.576 | 0.389 | 2.675 | 0.347 | 2.774 | 0.306 | 2.872 | 0.268 | 2.969 |
| 38 | 0.702 | 2.073 | 0.651 | 2.164 | 0.601 | 2.256 | 0.552 | 2.350 | 0.504 | 2.445 | 0.458 | 2.540 | 0.414 | 2.637 | 0.371 | 2.733 | 0.330 | 2.828 | 0.291 | 2.923 |
| 39 | 0.723 | 2.055 | 0.673 | 2.143 | 0.623 | 2.232 | 0.575 | 2.323 | 0.528 | 2.414 | 0.482 | 2.507 | 0.438 | 2.600 | 0.395 | 2.694 | 0.354 | 2.787 | 0.315 | 2.879 |
| 40 | 0.744 | 2.039 | 0.694 | 2.123 | 0.645 | 2.210 | 0.597 | 2.297 | 0.551 | 2.386 | 0.505 | 2.476 | 0.461 | 2.566 | 0.418 | 2.657 | 0.377 | 2.748 | 0.338 | 2.838 |
| 45 | 0.835 | 1.972 | 0.790 | 2.044 | 0.744 | 2.118 | 0.700 | 2.193 | 0.655 | 2.269 | 0.612 | 2.346 | 0.570 | 2.424 | 0.528 | 2.503 | 0.488 | 2.582 | 0.448 | 2.661 |
| 50 | 0.913 | 1.925 | 0.871 | 1.987 | 0.829 | 2.051 | 0.787 | 2.116 | 0.746 | 2.182 | 0.705 | 2.250 | 0.665 | 2.318 | 0.625 | 2.387 | 0.586 | 2.456 | 0.548 | 2.526 |
| 55 | 0.979 | 1.891 | 0.940 | 1.945 | 0.902 | 2.002 | 0.863 | 2.059 | 0.825 | 2.117 | 0.786 | 2.176 | 0.748 | 2.237 | 0.711 | 2.298 | 0.674 | 2.359 | 0.637 | 2.421 |
| 60 | 1.037 | 1.865 | 1.001 | 1.914 | 0.965 | 1.964 | 0.929 | 2.015 | 0.893 | 2.067 | 0.857 | 2.120 | 0.822 | 2.173 | 0.786 | 2.227 | 0.751 | 2.283 | 0.716 | 2.338 |
| 65 | 1.087 | 1.845 | 1.053 | 1.889 | 1.020 | 1.934 | 0.986 | 1.980 | 0.953 | 2.027 | 0.919 | 2.075 | 0.886 | 2.123 | 0.852 | 2.172 | 0.819 | 2.221 | 0.786 | 2.272 |
| 70 | 1.131 | 1.831 | 1.099 | 1.870 | 1.068 | 1.911 | 1.037 | 1.953 | 1.005 | 1.995 | 0.974 | 2.038 | 0.943 | 2.082 | 0.911 | 2.127 | 0.880 | 2.172 | 0.849 | 2.217 |
| 75 | 1.170 | 1.819 | 1.141 | 1.856 | 1.111 | 1.893 | 1.082 | 1.931 | 1.052 | 1.970 | 1.023 | 2.009 | 0.993 | 2.049 | 0.964 | 2.090 | 0.934 | 2.131 | 0.905 | 2.172 |
| 80 | 1.205 | 1.810 | 1.177 | 1.844 | 1.150 | 1.878 | 1.122 | 1.913 | 1.094 | 1.949 | 1.066 | 1.984 | 1.039 | 2.022 | 1.011 | 2.059 | 0.983 | 2.097 | 0.955 | 2.135 |
| 85 | 1.236 | 1.803 | 1.210 | 1.834 | 1.184 | 1.866 | 1.158 | 1.898 | 1.132 | 1.931 | 1.106 | 1.965 | 1.080 | 1.999 | 1.053 | 2.033 | 1.027 | 2.068 | 1.000 | 2.104 |
| 90 | 1.264 | 1.798 | 1.240 | 1.827 | 1.215 | 1.856 | 1.191 | 1.886 | 1.166 | 1.917 | 1.141 | 1.948 | 1.116 | 1.979 | 1.091 | 2.012 | 1.066 | 2.044 | 1.041 | 2.077 |
| 95 | 1.290 | 1.793 | 1.267 | 1.821 | 1.244 | 1.848 | 1.221 | 1.876 | 1.197 | 1.905 | 1.174 | 1.934 | 1.150 | 1.963 | 1.126 | 1.993 | 1.102 | 2.023 | 1.079 | 2.054 |
| 100 | 1.314 | 1.790 | 1.292 | 1.816 | 1.270 | 1.841 | 1.248 | 1.868 | 1.225 | 1.895 | 1.203 | 1.922 | 1.181 | 1.949 | 1.158 | 1.977 | 1.136 | 2.006 | 1.113 | 2.034 |
| 150 | 1.473 | 1.783 | 1.458 | 1.799 | 1.444 | 1.814 | 1.429 | 1.830 | 1.414 | 1.847 | 1.400 | 1.863 | 1.385 | 1.880 | 1.370 | 1.897 | 1.355 | 1.913 | 1.340 | 1.931 |
| 200 | 1.561 | 1.791 | 1.550 | 1.801 | 1.539 | 1.813 | 1.528 | 1.824 | 1.518 | 1.836 | 1.507 | 1.847 | 1.495 | 1.860 | 1.484 | 1.871 | 1.474 | 1.883 | 1.462 | 1.896 |

*Nota:* $n$ = número de observaciones.
$k'$ = número de variables explicativas, excluyendo el término constante.

Fuente: Savin y White, op. cit., con autorización de la Sociedad Econométrica.

## TABLA D.6A Valores críticos de rachas en la prueba de rachas

| | $N_2$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_1$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 2 | | | | | | | | | | | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | | | | | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | | | | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 |
| 5 | | | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 |
| 6 | | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 |
| 7 | | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 6 |
| 8 | | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 |
| 9 | | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 |
| 10 | | 2 | 3 | 3 | 4 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 9 |
| 11 | | 2 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 9 |
| 12 | 2 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 |
| 13 | 2 | 2 | 3 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| 14 | 2 | 2 | 3 | 4 | 5 | 5 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 10 | 11 | 11 |
| 15 | 2 | 3 | 3 | 4 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 |
| 16 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 | 12 |
| 17 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 13 |
| 18 | 2 | 3 | 4 | 5 | 5 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 |
| 19 | 2 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 13 |
| 20 | 2 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 9 | 9 | 10 | 10 | 11 | 12 | 12 | 13 | 13 | 13 | 14 |

*Nota:* Las tablas D.6A y D.6B dan los valores críticos de *n* rachas para diversos valores de $N_1$ (símbolo +) y $N_2$ (símbolo −). Para una prueba de rachas de una muestra, cualquier valor de *n* igual o menor que el que aparece en la tabla D.6A, o igual o superior al que aparece en la tabla D.6B, es significativo en el nivel de 0.05.

Fuente: Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, Nueva York, 1956, tabla F, pp. 252-253. Siegel adaptó las tablas de la fuente original: Frieda S. Swed y C. Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives", *Annals of Mathematical Statistics*, vol. 14, 1943. Se reproducen con permiso de McGraw-Hill Book Company y de *Annals of Mathematical Statistics*.

## TABLA D.6B Valores críticos de rachas en la prueba de rachas

| | $N_2$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_1$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 2 | | | | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | | | | |
| 4 | | | | 9 | 9 | | | | | | | | | | | | | | |
| 5 | | | 9 | 10 | 10 | 11 | 11 | | | | | | | | | | | | |
| 6 | | | 9 | 10 | 11 | 12 | 12 | 13 | 13 | 13 | 13 | | | | | | | | |
| 7 | | | | 11 | 12 | 13 | 13 | 14 | 14 | 14 | 14 | 15 | 15 | 15 | | | | | |
| 8 | | | | 11 | 12 | 13 | 14 | 14 | 15 | 15 | 16 | 16 | 16 | 16 | 17 | 17 | 17 | 17 | 17 |
| 9 | | | | | 13 | 14 | 14 | 15 | 16 | 16 | 16 | 17 | 17 | 18 | 18 | 18 | 18 | 18 | 18 |
| 10 | | | | | 13 | 14 | 15 | 16 | 16 | 17 | 17 | 18 | 18 | 18 | 19 | 19 | 19 | 20 | 20 |
| 11 | | | | | 13 | 14 | 15 | 16 | 17 | 17 | 18 | 19 | 19 | 19 | 20 | 20 | 20 | 21 | 21 |
| 12 | | | | | 13 | 14 | 16 | 16 | 17 | 18 | 19 | 19 | 20 | 20 | 21 | 21 | 21 | 22 | 22 |
| 13 | | | | | | 15 | 16 | 17 | 18 | 19 | 19 | 20 | 20 | 21 | 21 | 22 | 22 | 23 | 23 |
| 14 | | | | | | 15 | 16 | 17 | 18 | 19 | 20 | 20 | 21 | 22 | 22 | 23 | 23 | 23 | 24 |
| 15 | | | | | | 15 | 16 | 18 | 18 | 19 | 20 | 21 | 22 | 22 | 23 | 23 | 24 | 24 | 25 |
| 16 | | | | | | | 17 | 18 | 19 | 20 | 21 | 21 | 22 | 23 | 23 | 24 | 25 | 25 | 25 |
| 17 | | | | | | | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 23 | 24 | 25 | 25 | 26 | 26 |
| 18 | | | | | | | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 25 | 26 | 26 | 27 |
| 19 | | | | | | | 17 | 18 | 20 | 21 | 22 | 23 | 23 | 24 | 25 | 26 | 26 | 27 | 27 |
| 20 | | | | | | | 17 | 18 | 20 | 21 | 22 | 23 | 24 | 25 | 25 | 26 | 27 | 27 | 28 |

**EJEMPLO 2**

En una sucesión de 30 observaciones consistentes en 20 signos + (= $N_1$) y en 10 signos − (= $N_2$), los valores críticos de las rachas en el nivel de significancia de 0.05 son 9 y 20, como indican las tablas D.6A y D.6B, respectivamente. Por consiguiente, si en una aplicación se encuentra que el número de rachas es igual o menor que 9 o igual o mayor que 20, se rechaza la hipótesis (en un nivel de significancia de 0.05) de que la secuencia observada es aleatoria.

**TABLA D.7 Valores críticos Dickey-Fuller $t$ (= $\tau$) a 1% y 5%, así como valores $F$ para pruebas de raíz unitaria**

| Tamaño de la muestra | $t_{nc}$* | | $t_c$* | | $t_{ct}$* | | $F^{\dagger}$ | | $F^{\ddagger}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% | 1% | 5% |
| 25 | −2.66 | −1.95 | −3.75 | −3.00 | −4.38 | −3.60 | 10.61 | 7.24 | 8.21 | 5.68 |
| 50 | −2.62 | −1.95 | −3.58 | −2.93 | −4.15 | −3.50 | 9.31 | 6.73 | 7.02 | 5.13 |
| 100 | −2.60 | −1.95 | −3.51 | −2.89 | −4.04 | −3.45 | 8.73 | 6.49 | 6.50 | 4.88 |
| 250 | −2.58 | −1.95 | −3.46 | −2.88 | −3.99 | −3.43 | 8.43 | 6.34 | 6.22 | 4.75 |
| 500 | −2.58 | −1.95 | −3.44 | −2.87 | −3.98 | −3.42 | 8.34 | 6.30 | 6.15 | 4.71 |
| $\infty$ | −2.58 | −1.95 | −3.43 | −2.86 | −3.96 | −3.41 | 8.27 | 6.25 | 6.09 | 4.68 |

* Los subíndices nc, $c$ y tc denotan, respectivamente, que no hay término constante, que hay término constante y que está presente un término constante y uno de tendencia en la regresión (21.9.5).

† Los valores críticos de $F$ son para las hipótesis conjuntas en las que los términos constante y $\delta$ en (21.9.5) son al mismo tiempo cero.

‡ Los valores críticos de $F$ son para las hipótesis conjuntas en las cuales los términos constante, de tendencia y $\delta$ en (21.9.5) son al mismo tiempo iguales a cero.

Fuente: Adaptado de W.A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, Nueva York, 1976, p. 373 (para la prueba $\tau$), y D.A. Dickey y W.A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, vol. 49, 1981, p. 1063.

Apéndice E

# Resultados de computadora de EViews, MINITAB, Excel y STATA

En este apéndice presentamos los resultados de computadora con *EViews,* MINITAB, Excel y STATA, algunos paquetes estadísticos populares para regresión y rutinas estadísticas relacionadas. Se emplean los datos de la tabla E.1 del sitio Web del libro para ilustrar los resultados que arrojan estos paquetes. La tabla E.1 proporciona datos sobre la tasa de participación en la fuerza laboral civil (CLFPR), la tasa de desempleo civil (CUNR) y el salario promedio real por hora en dólares de 1982 (AHE82) para la economía de Estados Unidos de 1980 a 2002.

Aunque en muchos aspectos los resultados básicos de la regresión son parecidos en todos estos paquetes, hay diferencias de forma. Algunos paquetes presentan resultados de varios dígitos, mientras que otros los aproximan a cuatro o cinco dígitos. Algunos paquetes proporcionan directamente las tablas del análisis de varianza (ANOVA), en tanto que en otros es necesario derivarlas. También existen diferencias en algunos resúmenes estadísticos de los diversos paquetes. Escapa al propósito de este apéndice enumerar todas las diferencias en estos paquetes estadísticos. Recomendamos al lector consultar información en los sitios Web de estos paquetes.

## E.1 *EViews*

Con la versión 6 de *EViews* efectuamos la regresión de CLFPR sobre CUNR y AHE82 y obtuvimos los resultados de la figura E.1.

Es el formato normal en que se presentan los resultados de *EViews.* La primera parte de esta figura presenta los coeficientes de regresión, los errores estándar estimados, los valores *t* según la hipótesis nula de que los valores poblacionales correspondientes de estos coeficientes son cero, y los valores *p* de estos valores *t.* Enseguida se presentan $R^2$ y $R^2$ ajustada. El otro resumen en la primera parte relaciona el error estándar de la regresión, la suma de cuadrados residual (SCR) y el valor *F* para probar la hipótesis de que los (verdaderos) valores de todos los coeficientes de pendiente son, simultáneamente, iguales a cero. A menudo se usan los criterios de información de Akaike y Schwarz para elegir entre modelos rivales. Cuanto más bajo sea el valor de estos criterios, mejor será el modelo. El método de máxima verosimilitud (MV) es una alternativa al método de mínimos cuadrados. Así como en MCO se encuentran los estimadores que reducen la

**FIGURA E.1**
Resultados de *EViews* de la regresión de la participación en la fuerza laboral civil.

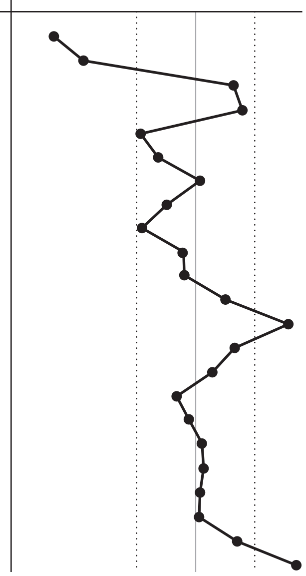
Variable dependiente: CLFPR
Método: Mínimos cuadrados
Muestra: 1980-2002
Observaciones incluidas: 23

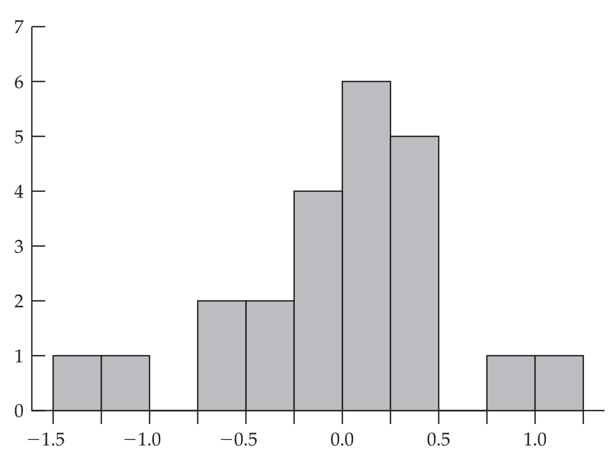
| Variable | Coeficiente | Error estándar | Estadístico *t* | Prob. |
|---|---|---|---|---|
| C | 80.90133 | 4.756195 | 17.00967 | 0.0000 |
| CUNR | −0.671348 | 0.082720 | −8.115928 | 0.0000 |
| AHE82 | −1.404244 | 0.608615 | −2.307278 | 0.0319 |

| | | | |
|---|---|---|---|
| R cuadrada | 0.772765 | Media de la variable dependiente | 65.89565 |
| R cuadrada ajustada | 0.750042 | Desviación estándar de la variable dependiente | 1.168713 |
| Error estándar de la regresión | 0.584308 | Criterio de información de Akaike | 1.884330 |
| Suma de cuadrados residual | 6.828312 | Criterio de Schwarz | 2.032438 |
| Log verosimilitud | −18.66979 | Estadístico *F* | 34.00731 |
| Estad. Durbin-Watson | 0.787625 | Prob. (estadístico *F*) | 0.000000 |

| Obs. | Real | Ajustado | Residuo |
|---|---|---|---|
| 1980 | 63.8000 | 65.2097 | −1.40974 |
| 1981 | 63.9000 | 65.0004 | −1.10044 |
| 1982 | 64.0000 | 63.6047 | 0.39535 |
| 1983 | 64.0000 | 63.5173 | 0.48268 |
| 1984 | 64.4000 | 64.9131 | −0.51311 |
| 1985 | 64.8000 | 65.1566 | −0.35664 |
| 1986 | 65.3000 | 65.2347 | 0.06526 |
| 1987 | 65.6000 | 65.8842 | −0.28416 |
| 1988 | 65.9000 | 66.4103 | −0.51027 |
| 1989 | 66.5000 | 66.6148 | −0.11476 |
| 1990 | 66.5000 | 66.5819 | −0.08186 |
| 1991 | 66.2000 | 65.8745 | 0.32546 |
| 1992 | 66.4000 | 65.4608 | 0.93923 |
| 1993 | 66.3000 | 65.8917 | 0.40834 |
| 1994 | 66.6000 | 66.4147 | 0.18530 |
| 1995 | 66.6000 | 66.7644 | −0.16441 |
| 1996 | 66.8000 | 66.8425 | −0.04251 |
| 1997 | 67.1000 | 67.0097 | 0.09032 |
| 1998 | 67.1000 | 66.9974 | 0.10263 |
| 1999 | 67.1000 | 67.0443 | 0.05569 |
| 2000 | 67.2000 | 67.1364 | 0.06355 |
| 2001 | 56.9000 | 66.4589 | 0.44105 |
| 2002 | 66.6000 | 65.5770 | 1.02304 |

**Gráfico de residuos**






Serie: Residuos
Muestra: 1980-2002
Observaciones: 23

| | |
|---|---|
| Media | −1.39e-14 |
| Mediana | 0.063552 |
| Máximo | 1.023040 |
| Mínimo | −1.409735 |
| Desv. est. | 0.557116 |
| Asimetría | −0.593013 |
| Curtosis | 3.752631 |
| Jarque-Bera | 1.890898 |
| Probabilidad | 0.388505 |

suma de cuadrados de los errores, en MV tratamos de encontrar los estimadores que maximizan la posibilidad de observar la muestra en cuestión. Según el supuesto de normalidad del término de error, MCO y MV proporcionan estimaciones idénticas de los coeficientes de regresión. Con el estadístico de Durbin-Watson se averigua si hay correlación serial de primer orden en los términos de error.

La segunda parte de los resultados de *EViews* proporciona los valores reales y ajustados de la variable dependiente y la diferencia entre los dos, que representa los residuos. La gráfica de estos residuos aparece a un lado de estos resultados con una línea vertical que denota cero. Los puntos a la derecha de la línea vertical son residuos positivos y los que aparecen a la izquierda residuos negativos.

La tercera parte de los resultados presenta el histograma de los residuos junto con un resumen estadístico. Da el estadístico de Jarque-Bera (JB) para probar la normalidad de los términos de error y también presenta la probabilidad de obtener los estadísticos indicados. Cuanto más alta sea la probabilidad de obtener el estadístico JB observado, mayor será la evidencia en favor de la hipótesis nula de que los términos de error están distribuidos normalmente.

Observe que *EViews* no proporciona directamente la tabla de análisis de varianza (ANOVA), pero se genera con facilidad a partir de los datos sobre la suma de cuadrados residual, la suma de cuadrados total (que a su vez debe derivarse de la desviación estándar de la variable dependiente) y los grados de libertad asociados. El valor $F$ obtenido en este ejercicio debe ser igual al valor $F$ reportado en la primera parte de la tabla.

## E.2 MINITAB

Con la versión 15 de MINITAB y los mismos datos obtuvimos los resultados de la regresión de la figura E.2.

MINITAB presenta primero la regresión múltiple estimada. Esto va seguido de una lista de variables predictoras (es decir, explicativas), los coeficientes de regresión estimados, los errores estándar, los valores $T$ $(= t)$ y los valores $p$. En estos resultados $S$ representa el error estándar de la estimación, y los valores de $R^2$ y $R^2$ ajustada se presentan en forma porcentual.

A continuación se presenta la tabla ANOVA habitual. Una característica de la tabla ANOVA es que desglosa la regresión, o suma de cuadrados explicada, entre las variables predictoras. Así, de la suma de cuadrados total de la regresión de 23.226, la parte correspondiente a CUNR es 21.404 y la que corresponde a AHE82 es 1.822, lo que indica que, en términos relativos, CUNR tiene un efecto más importante en CLFPR que AHE82.

Una característica exclusiva de los resultados de la regresión de MINITAB es que reportan las observaciones “atípicas”, es decir, las observaciones de algún modo diferentes del resto de las observaciones de la muestra. Se hace referencia a esto en la gráfica de residuos presentada en los resultados de *EViews,* pues muestra que las observaciones 1 y 23 se encuentran muy lejos de la línea de cero que ahí se presenta. MINITAB también produce una gráfica de residuos parecida a la de *EViews.* El término Resid est en estos resultados significa residuos estandarizados, es decir, los residuos divididos entre $S$, el error estándar de la estimación.

Al igual que *EViews,* MINITAB también presenta el estadístico de Durbin-Watson y el histograma de residuos. El histograma es una representación visual. Si su forma se parece a la distribución normal, es probable que los residuos estén distribuidos normalmente. La gráfica de probabilidad normal cumple el mismo propósito. Si los residuos estimados se encuentran aproximadamente en línea recta, podemos afirmar que están distribuidos normalmente. El estadístico de Anderson-Darling (AD), un estadístico asociado a la gráfica de la probabilidad normal, prueba la hipótesis de que la variable en consideración (en este caso, los residuos) está distribuida normalmente. Si el valor $p$ del estadístico AD calculado es razonablemente alto, por ejemplo, superior a 0.10, concluimos que la variable está distribuida normalmente. En este ejemplo, el estadístico AD tiene un valor de 0.481 con un valor $p$ de alrededor de 0.21 o 21%. Por tanto, la conclusión es que los residuos obtenidos del modelo de regresión tienen distribución normal.

**FIGURA E.2** Resultados de MINITAB de la tasa de participación en la fuerza laboral civil.

**Análisis de regresión: CLFPR sobre CUNR, AHE82**

La ecuación de regresión es:
CLFPR = 81.0 − 0.672 CUNR − 1.41 AHE82

| Predictor | Coef. | SE coef. | T | P |
|---|---|---|---|---|
| Constante | 80.951 | 4.770 | 16.97 | 0.000 |
| CUNR | −0.67163 | 0.08270 | −8.12 | 0.000 |
| AHE82 | −1.4104 | 0.6103 | −2.31 | 0.032 |

$S$ = 0.584117  $R$ cuadrada = 77.3%  $R$ cuadrada (ajust.) = 75%

Análisis de varianza

| Origen | GL | SC | SP | F | P |
|---|---|---|---|---|---|
| Regresión | 2 | 23.226 | 11.613 | 34.04 | 0.000 |
| Error residual | 20 | 6.824 | 0.341 | | |
| Total | 22 | 30.050 | | | |

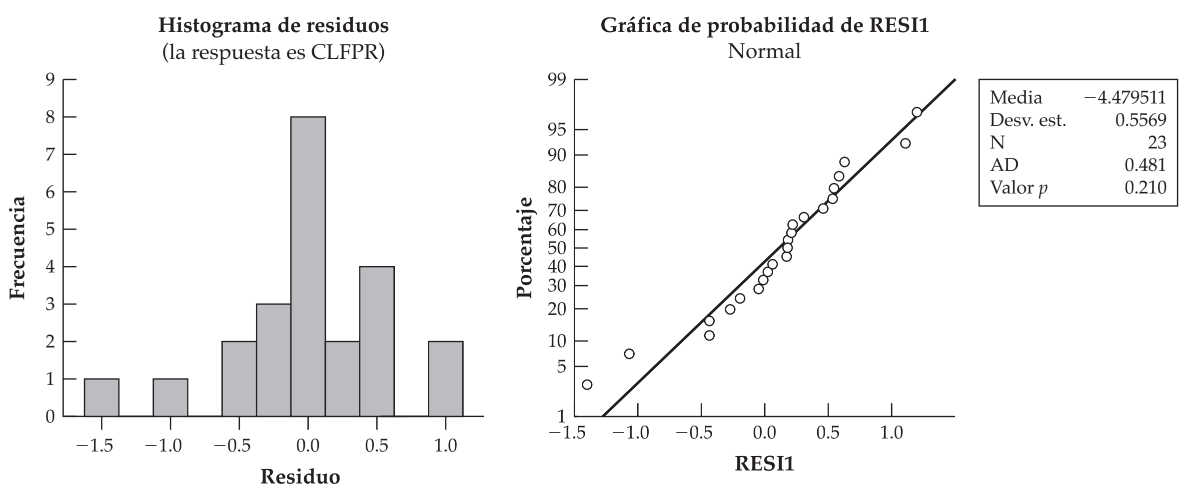
| Origen | GL | SC sec. |
|---|---|---|
| CUNR | 1 | 21.404 |
| AHE82 | 1 | 1.822 |

Observaciones atípicas

| Obs. | CUNR | CLFPR | Ajustado | ES ajustado | Residuo | Resid est. |
|---|---|---|---|---|---|---|
| 1 | 7.10 | 63.800 | 65.209 | 0.155 | −1.409 | −2.50*R* |
| 23 | 5.80 | 66.600 | 65.575 | 0.307 | 1.025 | 2.06*R* |

*R* denota una observación con un residuo estandarizado grande.

Estadístico de Durbin-Watson = 0. 787065




# E.3 Excel

Con Microsoft Excel obtuvimos los resultados de la regresión de la tabla E.2.

Excel presenta primero un resumen estadístico, como $R^2$, $R$ múltiple, que es la raíz cuadrada (positiva) de $R^2$, $R^2$ ajustada y el error estándar de la estimación. Luego presenta la tabla ANOVA. A continuación presenta los coeficientes estimados, los errores estándar, los valores $t$ de los co-

**TABLA E.2**
**Resultados de Excel de la tasa de participación en la fuerza laboral civil**

**Resumen de resultados**

**Estadísticas de la regresión**

| | |
|---|---|
| *R* múltiple | 0.879155 |
| *R* cuadrada | 0.772914 |
| *R* ajustada | 0.750205 |
| Error estándar | 0.584117 |
| Observación | 23 |

**ANOVA**

| | gl | SC | SP | F | Significancia de *F* |
|---|---|---|---|---|---|
| Regresión | 2 | 23.22572 | 11.61286 | 34.03611 | 3.65E-07 |
| Residuo | 20 | 6.823846 | 0.341192 | | |
| Total | 22 | 30.04957 | | | |

| | Coeficiente | Error estándar | Est. *t* | Valor *p* | 95% inferior | 95% superior |
|---|---|---|---|---|---|---|
| Intercept | 80.95122 | 4.770337 | 16.96971 | 2.42E-13 | 71.00047 | 90.90196 |
| CUNR | −0.671631 | 0.082705 | −8.120845 | 9.24E-08 | −0.84415 | −0.499112 |
| AHE82 | −1.410432 | 0.610348 | −2.310867 | 0.031626 | −2.683594 | −0.13727 |

eficientes estimados y sus valores *p*. También proporciona los valores reales y estimados de la variable dependiente y la gráfica de residuos, así como la gráfica de probabilidad normal.

Una característica única de Excel es que proporciona el intervalo de confianza a 95% (o cualquier porcentaje especificado) de los verdaderos valores de los coeficientes estimados. Así, el valor estimado del coeficiente de CUNR es −0.671631 y el intervalo de confianza del verdadero valor del coeficiente de CUNR es (−0.84415 a −0.499112). Esta información es muy valiosa para las pruebas de hipótesis.

## E.4 STATA

Con STATA obtuvimos los resultados de la regresión de la tabla E.3.

STATA presenta primero la tabla del análisis de varianza junto con un resumen de estadísticos, como $R^2$, $R^2$ ajustada y la raíz del error cuadrático medio (ECM), que es simplemente el error estándar de la regresión.

Enseguida proporciona los valores de los coeficientes estimados, sus errores estándar y valores *t*, y los valores *p* de los estadísticos *t,* así como el intervalo de confianza a 95% de cada coeficiente de regresión, de forma similar a los resultados de Excel.

## E.5 Comentarios finales

Presentamos sólo los resultados básicos de estos paquetes para ilustrar el ejemplo. Sin embargo, cabe señalar que los paquetes como *EViews* y STATA son muy amplios y contienen muchas de las técnicas econométricas que estudiamos en este libro. Una vez que aprende a entrar a estos paquetes, la ejecución de las diversas subrutinas es cuestión de práctica. Si desea seguir estudiando econometría, es conveniente que adquiera uno o más de estos paquetes.

**TABLA E.3**
**Resultados de STATA de la tasa de participación en la fuerza laboral civil**

Estadística/análisis de datos
Proyecto: Datos de la tabla E.1

8.0
Estadística/análisis de datos

Copyright 1984–2003
Stata Corporation
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

regresión de clfpr sobre cunr ahe82

| Origen | SC | gl | SP |
|---|---|---|---|
| Modelo | 23.2256929 | 2 | 11.6128465 |
| Residuo | 6.82384072 | 20 | .341192036 |
| Total | 30.0495337 | 22 | 1.36588789 |

Número de obs. = 23
F(2, 20) = 34.04
Prob > F = 0.0000
R cuadrada = 0.7729
R cuadrada ajust. = 0.7502
Raíz de ECM = .58412

| clfpr | Coef. | Error estándar | *t* | *p* > \|*t*\| | [Intervalo de conf. a 95%] | |
|---|---|---|---|---|---|---|
| cunr | −.6716305 | .0827045 | −8.12 | 0.000 | −.8441491 | −.4991119 |
| ahe82 | −1.410433 | .6103473 | −2.31 | 0.032 | −2.683595 | −.1372707 |
| _cons | 80.95122 | 4.770334 | 16.97 | 0.000 | 71.00048 | 90.90197 |

# Referencias

www.eviews.com
www.stata.com
www.minitab.com
Microsoft Excel

R. Carter Hill, William E. Griffiths y George G. Judge, *Using Excel for Undergraduate Econometrics,* John Wiley & Sons, Nueva York, 2001.

Apéndice F

# Datos económicos en la World Wide Web*

*Economic Statistics Briefing Room:* Es una fuente excelente de datos sobre producto, ingreso, empleo, desempleo, utilidades, producción y actividades empresariales, así como de precios y dinero, créditos y mercados de valores, y estadísticas internacionales.

**http://www.whitehouse.gov/fsbr/esbr.htm**

*Federal Reserve System Beige Book:* Proporciona un resumen de las condiciones actuales de la economía por cada distrito de la Reserva Federal. Actualmente existen 12 distritos de la Reserva Federal.

**http://www.federalreserve.gov/FOM/BEIGEBOOK**

*Página del National Bureau of Economic Research (NBER):* Este instituto privado de investigación económica, que goza de mucho prestigio y reconocimiento, tiene una gran cantidad de datos sobre precios de activos, mano de obra, productividad, oferta de dinero, indicadores de ciclos económicos, etc. NBER proporciona muchos enlaces con otros sitios Web.

**http://www.nber.org**

*Panel Study:* Proporciona datos de encuestas longitudinales sobre muestras representativas de individuos y familias estadounidenses. Estos datos se recopilan anualmente desde 1968.

**http://psidonline.isr.umich.edu/**

*Resources for Economists on the Internet:* Fuente muy completa de información y datos sobre muchas actividades económicas que tiene enlaces con gran cantidad de sitios Web. Es una fuente muy valiosa para los economistas académicos y no académicos.

**http://rfe.org/**

*American Stock Exchange:* Dispone de información sobre las casi 700 compañías que cotizan en el segundo mercado bursátil más grande del mundo.

**http://www.amex.com/**

*Página del Bureau of Economic Analysis (BEA):* Esta oficina del Departamento de Comercio de Estados Unidos, que publica *Survey of Current Business*, es una excelente fuente de datos sobre todo tipo de actividades económicas.

**http://www.bea.gov/**

*Publicaciones de la CIA:* En esta página se puede consultar el *World Fact Book* (anuario) y el *Handbook of International Statistics*.

**http://www.cia.gov/library/publications**

* Adaptado de *Annual Editions: Microeconomics 98/99*, ed. Don Cole, Dushkin/McGraw-Hill, Connecticut, 1998. Debe señalarse que esta lista no es exhaustiva. Las fuentes aquí incluidas se actualizan continuamente.

*Energy Information Administration (DOE):* Contiene información económica y datos sobre cada categoría de combustibles.

**http://www.eia.doe.gov/**

*FRED Database:* El Banco de la Reserva Federal de St. Louis publica datos históricos económicos y sociales en los que se incluyen tasas de interés, indicadores monetarios y de negocios, tipos de cambio, etcétera.

**http://research.stlouisfed.org/fred2/**

*International Trade Administration:* Ofrece muchos enlaces Web con estadísticas sobre comercio, programas internacionales, etcétera.

**http://trade.gov/index.asp**

*STAT-USA Databases:* El National Trade Data Bank representa la fuente más completa de datos sobre comercio internacional e información sobre fomento a la exportación. Cuenta con una gran cantidad de datos sobre condiciones demográficas, políticas y socioeconómicas en distintos países.

**http://www.stat-usa.gov/**

*Statistical Resources on the Web/Economics:* Es una excelente fuente de datos estadísticos cotejados de diversas dependencias federales estadounidenses, indicadores económicos, la Junta de la Reserva Federal, datos sobre precios al consumidor y enlaces Web con otras fuentes.

**http://www.lib.umich.edu/govdocs/stats.html**

*Bureau of Labor Statistics:* La página principal contiene datos relacionados con diversos aspectos del empleo, desempleo y salarios; proporciona también enlaces con otros sitios Web de estadísticas.

**http://www.stats.bls.gov/**

*Página de U.S. Census Bureau:* Es la fuente primordial de datos sociales, demográficos y económicos sobre ingreso, empleo, distribución del ingreso y pobreza.

**http://www.census.gov/**

*General Social Survey:* Encuesta anual mediante entrevistas personales sobre las familias estadounidenses, la cual empezó a realizarse en 1972. Más de 35 000 familias han respondido a unas 2 500 preguntas que abarcan una gran variedad de datos.

**http://www.norc.org/GSS+website/**

*Institute for Research on Poverty:* Datos recopilados por un centro de investigación universitario, no lucrativo y apolítico, respecto de la pobreza y la desigualdad social.

**http://www.irp.wisc.edu/**

*Social Security Administration:* Sitio web oficial de la Social Security Administration que contiene una gran variedad de datos.

**http://www.ssa.gov/**

# Bibliografía selecta


## Introductoria

Frank, C. R., Jr., *Statistics and Econometrics*, Holt, Rinehart and Winston, Nueva York, 1971.

Goldberger, Arthur S., *Introductory Econometrics*, Harvard University Press, 1998.

Gujarati, Damodar N., *Essentials of Econometrics*, 3a. ed., McGraw-Hill, Nueva York, 2006.

Halcoussis, Dennis, *Understanding Econometrics,* Thomson, 2005.

Hill, Carter, William Griffiths y George Judge, *Undergraduate Econometrics*, John Wiley & Sons, Nueva York, 2001.

Hu, Teh-Wei, *Econometrics, An Introductory Analysis*, University Park Press, Baltimore, 1973.

Katz, David A., *Econometric Theory and Applications*, Prentice Hall, Englewood Cliffs, N.J., 1982.

Klein, Lawrence R., *An Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1962.

Koop, Gary, *Analysis of Economic Data*, John Wiley & Sons, Nueva York, 2000.

Schmidt, Stephen, J., *Econometrics,* McGraw-Hill, Nueva York, 2005.

Walters, A. A., *An Introduction to Econometrics*, Macmillan, Londres, 1968.

## Intermedia

Aigner, D. J., *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1971.

Dhrymes, Phoebus J., *Introductory Econometrics*, Springer-Verlag, Nueva York, 1978.

Dielman, Terry E., *Applied Regression Analysis for Business and Economics*, PWS-Kent, Boston, 1991.

Dougherty, Christopher, *Introduction to Econometrics,* 3a. ed., Oxford University Press, Oxford, 2007.

Draper, N. R. y H. Smith, *Applied Regression Analysis*, 3a. ed., John Wiley & Sons, Nueva York, 1998.

Dutta, M., *Econometric Methods*, South-Western Publishing Company, Cincinnati, 1975.

Goldberger, A. S., *Topics in Regression Analysis*, Macmillan, Nueva York, 1968.

Griffiths, William E., R. Carter Hill y George G. Judge, *Learning and Practicing Econometrics*, John Wiley & Sons, Nueva York, 1993.

Harris, Richard y Robert Sollis, *Applied Time Series Modelling and Forecasting,* John Wiley & Sons, Inglaterra, 2003.

Heij, Christiaan, Paul deBoer, Philip Hans Franses, Teun Kloek y Herman K. van Djik, *Econometric Methods with Applications in Business and Economics,* Oxford University Press, Nueva York, 2004.

Huang, D. S., *Regression and Econometric Methods*, John Wiley & Sons, Nueva York, 1970.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl y Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, Nueva York, 1982.

Kelejian, H. A. y W. E. Oates, *Introduction to Econometrics: Principles and Applications*, 2a. ed., Harper & Row, Nueva York, 1981.

Koutsoyiannis, A., *Theory of Econometrics*, Harper & Row, Nueva York, 1973.

Maddala, G. S., *Introduction to Econometrics*, 3ª ed., John Wiley & Sons, Nueva York, 2001.

Mark, Stewart B. y Kenneth F. Wallis, *Introductory Econometrics*, 2a. ed., John Wiley & Sons, Nueva York, 1981. Un libro de Halsted Press.

Murphy, James L., *Introductory Econometrics*, Richard D. Irwin, Homewood, Ill., 1973.

Nachane, Dilip M., *Econometrics: Theoretical Foundations and Empirical Perspectives,* Oxford University Press, Nueva Delhi, 2006.

Netter, J. y W. Wasserman, *Applied Linear Statistical Models*, Richard D. Irwin, Homewood, Ill., 1974.

Pindyck, R. S. y D. L. Rubinfeld, *Econometric Models and Econometric Forecasts*, 4a. ed., McGraw-Hill, Nueva York, 1990.

Sprent, Peter, *Models in Regression and Related Topics*, Methuen, Londres, 1969.

Stock, James H. y Mark W. Watson, *Introduction to Econometrics,* 2a. ed., Pearson/Addison-Wesley, Boston, 2007.

Tintner, Gerhard, *Econometrics*, John Wiley & Sons (science ed.), Nueva York, 1965.

Valavanis, Stefan, *Econometrics: An Introduction to Maximum-Likelihood Methods*, McGraw-Hill, Nueva York, 1959.

Verbeek, Marno, *A Guide to Modern Econometrics*, John Wiley & Sons, Nueva York, 2000.

Wonnacott, R. J. y T. H. Wonnacott, *Econometrics*, 2a. ed., John Wiley & Sons, Nueva York, 1979.

Wooldridge, Jeffrey M., *Introductory Econometrics*, 3a. ed., South-Western College Publishing, 2000.

## Avanzada

Cameron, A. Colin y Pravin K. Trivedi, *Microeconomics: Methods and Applications,* Cambridge University Press, Nueva York, 2005.

Chow, Gregory C., *Econometric Methods*, McGraw-Hill, Nueva York, 1983.

Christ, C. F, *Econometric Models and Methods*, John Wiley & Sons, Nueva York, 1966.

Davidson, James, *Econometric Theory*, Blackwell Publishers, Oxford, Inglaterra, 2000.

Dhrymes, P. J., *Econometrics: Statistical Foundations and Applications*, Harper & Row, Nueva York, 1970.

Fomby, Thomas B., Carter R. Hill y Stanley R. Johnson, *Advanced Econometric Methods*, Springer-Verlag, Nueva York, 1984.

Goldberger, A. S., *Econometric Theory*, John Wiley & Sons, Nueva York, 1964.

——————, *A Course in Econometrics*, Harvard University Press, Cambridge, Mass., 1991.

Greene, William H., *Econometric Analysis*, 4a. ed., Prentice Hall, Englewood Cliffs, N. J., 2000.

Harvey, A. C., *The Econometric Analysis of Time Series*, 2a. ed., MIT Press, Cambridge, Mass., 1990.

Hayashi, Fumio, *Econometrics*, Princeton University Press, Princeton, N. J., 2000.

Johnston, J., *Econometric Methods*, 3a. ed., McGraw-Hill, Nueva York, 1984.

Judge, George G., Carter R. Hill, William E. Griffiths, Helmut Lütkepohl y Tsoung-Chao Lee, *Theory and Practice of Econometrics*, John Wiley & Sons, Nueva York, 1980.

Klein, Lawrence R., *A Textbook of Econometrics*, 2a. ed., Prentice Hall, Englewood Cliffs, N.J., 1974.

Kmenta, Jan, *Elements of Econometrics*, 2a. ed., Macmillan, Nueva York, 1986.

Madansky, A., *Foundations of Econometrics*, North-Holland, Ámsterdam, 1976.

Maddala, G. S., *Econometrics*, McGraw-Hill, Nueva York, 1977.

Malinvaud, E., *Statistical Methods of Econometrics*, 2a. ed., North-Holland, Ámsterdam, 1976.

Mills, Terence C. y Kerry Patterson, *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory,* Palgrave/Macmillan, Nueva York, 2006.

Mittelhammer, Ron C., George G. Judge y Douglas J. Miller, *Econometric Foundations*, Cambridge University Press, Nueva York, 2000.

Peracchi, Franco, *Econometrics,* John Wiley & Sons, Nueva York, 2001.

Theil, Henry, *Principles of Econometrics*, John Wiley & Sons, Nueva York, 1971.

## Especializada

Belsley, David A., Edwin Kuh y Roy E. Welsh, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, Nueva York, 1980.

Dhrymes, P. J., *Distributed Lags: Problems of Estimation and Formulation*, Holden-Day, San Francisco, 1971.

Diebold, Francis X., *Elements of Forecasting*, 2a. ed., South-Western Publishers, 2001.

Goldfeld, S. M. y R. E. Quandt, *Nonlinear Methods of Econometrics*, North-Holland, Ámsterdam, 1972.

Gourieroux, Christian, *Econometrics of Qualitative Dependent Variables*, Cambridge University Press, Nueva York, 2000.

Graybill, F. A., *An Introduction to Linear Statistical Models*, vol. 1, McGraw-Hill, Nueva York, 1961.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, Princeton, N.J., 1994.

Maddala, G. S. y Kim In-Moo, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, Nueva York, 1998.

Mills, T. C., *Time Series Techniques for Economists*, Cambridge University Press, 1990.

Rao, C. R., *Linear Statistical Inference and Its Applications*, 2a. ed., John Wiley & Sons, Nueva York, 1975.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, Nueva York, 1971.

## Aplicada

Berndt, Ernst R., *The Practice of Econometrics*, Classic and Contemporary, Addison-Wesley, 1991.

Bridge, J. I., *Applied Econometrics*, North-Holland, Amsterdam, 1971.

Charemza, Wojciech W. y Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, 2a. ed., Edward Elgar Publisher, Nueva York, 1997.

Cramer, J. S., *Empirical Econometrics*, North-Holland, Amsterdam, 1969.

Desai, Meghnad, *Applied Econometrics*, McGraw-Hill, Nueva York, 1976.

Kennedy, Peter, *A Guide to Econometrics*, 4a. ed., MIT Press, Cambridge, Mass., 1998.

Leser, C. E. V., *Econometric Techniques and Problems*, 2a. ed., Hafner, Londres, 1974.

Mills, T. C., *The Econometric Modelling of Financial Time Series*, Cambridge University Press, 1993.

Mukherjee, Chandan, Howard White y Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, Nueva York, 1998.

Patterson, Kerry, *An Introduction to Applied Econometrics*, *A Time Series Approach*, St. Martin's Press, Nueva York, 2000.

Rao, Potluri y Roger LeRoy Miller, *Applied Econometrics*, Wadsworth, Belmont, Calif., 1971.

*Nota:* Para una lista de los artículos fundamentales sobre los diversos temas analizados en este libro, favor de consultar la extensa bibliografía dada al final de cada capítulo en Fomby *et al.*, citado arriba.

# Índice de nombres

Los números de página seguidos por una n indican material contenido en las notas.

# Índice analítico

Los números de página seguidos por una n indican material contenido en las notas.

D

E

F

G



H



I

## R



## S